Published in final edited form as:
Comput Human Behav. 2009 May 1; 25(3): 695–710.

Too real for comfort? Uncanny responses to computer generated faces

Karl F. MacDorman[*], **Robert D. Green**, **Chin-Chang Ho**, and **Clinton T. Koch**
Indiana University School of Informatics, IT 487, 535 West Michigan Street, Indianapolis, IN 46202, USA

Abstract

As virtual humans approach photorealistic perfection, they risk making real humans uncomfortable. This intriguing phenomenon, known as the *uncanny valley*, is well known but not well understood. In an effort to demystify the causes of the uncanny valley, this paper proposes several perceptual, cognitive, and social mechanisms that have already helped address riddles like empathy, mate selection, threat avoidance, cognitive dissonance, and psychological defenses. In the four studies described herein, a computer generated human character's facial proportions, skin texture, and level of detail were varied to examine their effect on perceived eeriness, human likeness, and attractiveness. In Study I, texture photorealism and polygon count increased human likeness. In Study II, texture photorealism heightened the accuracy of human judgments of ideal facial proportions. In Study III, atypical facial proportions were shown to be more disturbing on photorealistic faces than on other faces. In Study IV, a mismatch in the size and texture of the eyes and face was especially prone to make a character eerie. These results contest the depiction of the uncanny valley as a simple relation between comfort level and human likeness. This paper concludes by introducing a set of design principles for bridging the uncanny valley.

Keywords

Anthropomorphism; Facial perception; Masahiro Mori; Social cognitive neuroscience; Uncanny valley

1. Introduction

1.1. The bleeding edge of human photorealism

Computer graphics (CG) characters are challenging our ability to discern what is human. For example, the CG character Davy Jones looked so human in *Pirates of the Caribbean: At World's End* film critics assumed he was portrayed by an actor wearing prosthetic tentacles (Zacharek, 2007). The critics did not realize the actor, Bill Nighy, had been entirely replaced by digital artistry. Nevertheless, this digital Davy Jones was believable, because he was meant to look supernatural and creepy. The same principle applies to other CG villains, such as Gollum in *The Lord of the Rings* trilogy. By contrast, characters designed to look like real

[*]Corresponding author. Tel.: +1 317 215 7040; fax: +1 206 350 6089. *URL:*http://www.macdorman.com (K.F. MacDorman).

people have been less convincing, such as the CG heroes in *The Polar Express* and *Final Fantasy: The Spirits Within* (Geller, 2008; Pollick, in press). Thus, the achievement of photorealistic human character animation has remained elusive despite its status as a holy grail of computer graphics (MacGillivray, 2007).

The difficulty of human photorealism has been attributed to the uncanny valley (*bukimi no tani* in Japanese). The term derives from a hypothetical graph proposed in 1970 by Masahiro Mori (Fig. 1). The graph predicts that as something looks more human it also looks more agreeable, until it comes to look so human we start to find its nonhuman imperfections unsettling (MacDorman & Ishiguro, 2006; Mori, 1970). The imperfections expose a mismatch between the human qualities we are led to expect and the nonhuman qualities that instead follow—or vice versa. As examples of things that lie in the uncanny valley, Mori cites corpses, zombies, mannequins coming to life, and lifelike prosthetic hands (Mori, 1970). According to Mori, on a dark night a woman could mistake a prosthetic hand for a real one. If she then shook hands with it, upon feeling its coldness and hardness she might shriek with horror.

Concerns about the uncanny valley have taken on new urgency with the ever increasing use of CG animation. These concerns are often reported in trade journals and the popular press because of the uncanny valley's perceived impact on the multi-billion dollar animation and video game industries (Gouskos, 2006; MacMillan, 2007). The uncanny valley has even led studios like Pixar to shy away from human photorealism, choosing instead cartoony stylization (e.g., the characters of *The Incredibles*; Canemaker, 2004). The uncanny valley especially worries video game designers, because their animations are rendered instantaneously, without time for careful staging or touching up.

This paper's goal is to take a few small steps toward bridging the uncanny valley by uncovering some of its causes and proposing design principles to help photorealistic human characters escape from the valley. The eventual fulfillment of this goal could have an enormous economic impact and change the course of computer graphics animation and video games.

1.2. Possible explanations of the uncanny valley

Why would human beings be put-off by nonhuman features in a human-looking character when they feel unperturbed by the same features in a more stylized character? Possible answers found in the literature (MacDorman & Ishiguro, 2006; MacDorman, Vasudevan, & Ho, 2009) may be divided into two groups: those that involve automatic, stimulus-driven, specialized processing that occurs early in perception and those that involve a broader and more general range of cognitive processing that occurs later. Both kinds of processing engage affective and motor processing and are simultaneously active in perceiving human-looking forms. Although their behavioral components are the historical domain of perceptual and social psychology, respectively, they are now open to exploration through brain imaging. They may be separated through experimental procedures, such as the subliminal and supraliminal presentation of stimuli during the measurement of event-related potentials in the brain (Del Cul, Baillet, & Dehaene, 2007).

1.2.1. Explanations involving specialized perceptual processing—Recognition deficits caused by brain injury (e.g., prosopagnosia; Farah, Rabinowitz, Quinn, & Liu, 2000), face inversion and configurational effects, and the results of brain imaging studies and singleneuron studies in nonhuman primates indicate that face recognition is anatomically and functionally specialized (Carmel & Bentin, 2002), involving as many as six regions in the ventral visual pathway (Barbeau et al., 2008). Brain imaging has revealed that the fusiform face area (FFA) of the ventral occipito-temporal cortex responds with high selectivity to faces (Kanwisher, McDermott, & Chun, 1997). The FFA has been shown to be more active when the participant sees the stimulus as a face during the bistable oscillation of the Rubin face-vase illusion (Andrews, Schluppeck, Homfray, Matthews, & Blakemore, 2002; Hasson, Hendler, Ben Bashat, & Malach, 2001) or in near-threshold images (Grill-Spector, Knouf, & Kanwisher, 2004). Inverted presentation more greatly hinders a person's recall of faces than of other objects, except in domains of exceptional expertise (e.g., the recall of show dogs by an highly experienced judge; Diamond & Carey, 1986). Nevertheless, the degree of brain specialization for face perception is still contested (Downing, Jiang, Shuman, & Kanwisher, 2001; Gauthier & Logothetis, 2000; Haxby et al., 2001; Spiridon & Kanwisher, 2002).

1.2.1.1. Threat avoidance: Mori (1970) suspected the uncanny valley arose from the need for self-preservation. Christian Keysers elaborated this view from an evolutionary perspective, drawing on Rozin's theory of disgust (MacDorman & Ishiguro, 2006;Rozin & Fallon, 1987). Keysers posited that the uncanny valley is the result of an evolved mechanism for pathogen avoidance. The more human an organism looks, the stronger the aversion to its defects, because (1) defects indicate disease, (2) more human-looking organisms are more closely related to human beings genetically, and (3) the probability of contracting disease-causing bacteria, viruses, and other parasites increases with genetic similarity. Thus, leprosy looks disgusting to us, but leaf spot does not. A mechanism for pathogen avoidance would explain the strong tendency to be more sensitive to defects in our own species–and to defects in CG human characters and other human-looking entities–than to defects in distantly related species.

Perceived defects in a human-looking entity could trigger an aversive response automatically by activating an evolved mechanism for self-preservation. For example, according to Rozin and Fallon (1987), disgust originated in the distaste system of our earliest ancestors: the system that elicits mouth gape (Ekman & Friesen, 1986), revulsion, and nausea when something tastes bitter or rotten. However, the elicitors of disgust broadened from taste to include other senses like smell that are able to indicate objects at a distance that should be avoided. For human beings the elicitors broadened further to include nonperceptual inferences—not just how something is perceived but what it is interpreted as signifying. Thus, the realization that someone has committed a moral transgression (e.g., adultery) may also access the distaste system through preadaptation (Rozin & Fallon, 1987). Among the strongest elicitors of disgust are reminders of the animal nature of human beings and especially their mortality (Haidt, McCauley, & Rozin, 1993). Perceived defects in a human-looking entity can also elicit fear-motivated aversion by triggering a fight-or-flight response (Ohman, 2000). Fear is commonly associated with activation in the amygdala but may occur in its absence (Atkinson, Heberlein, & Adolphs, 2007; Davis & Whalen, 2001).

1.2.1.2. Shared circuits for empathy: Some research indicates that perceptual, cognitive, and affective processing may work in concert during the perception of uncanny forms (Chaminade, Hodgin, & Kawato, 2007; Krach et al., 2008). These shared circuits in the brain are thought to support the ability to understand the intentions of others, because they are active both when someone performs an intentional action and when that person sees someone else perform the same action (Keysers & Gazzola, 2007). A pair of studies have shown that human-looking entities activate these shared circuits more powerfully (Chaminade et al., 2007; Krach et al., 2008), while other studies have found no significant difference between human beings and robots (Gazzola, Rizzolatti, Wicker, & Keysers, 2007). Using functional magnetic resonance imaging (fMRI) of human brains, Krach et al. (2008) discovered a linear relation between a robot's human likeness and cortical activation in the medial frontal cortex and the right temporoparietal junction. The shared circuits implicated in "mentalizing" about others' intentions have been referred to as the *mirror system* and identified with the premotor cortex in macaque monkeys (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and human beings (Grezes, Armony, Rowe, & Passingham, 2003). In a study using positron emission tomography (PET) scans, Tai, Scherfler, Brooks, Sawamoto, and Castiello (2004) found a significant neural response in the left premotor cortex when human participants watched an action being performed by a human being but not when performed by a robot. These brain imaging results complement behavioral evidence, which indicates that both the feeling of being understood and the frequency of soindustrial cial responses increases with the human likeness of a computer interface or virtual agent (Burgoon et al., 2000; Gong, 2008).

While the ability to understand the intentions of others is a key element of empathy, other shared circuits enable us to experience the emotions of others more directly (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Preston & de Waal, 2002). Jabbi, Bastiaansen, and Keysers (2008) found that experiencing disgust, viewing someone else experiencing it, or imagining a disgusting experience have a common neural substrate in the anterior insular cortex and adjacent frontal operculum. Human appearance and emotional expressivity may heighten empathy by enhancing the brain's ability to simulate being in another person's place (Cole, 2001; Preston & de Waal, 2002).

1.2.1.3. Evolutionary aesthetics: People judge the attractiveness of others at a glance; they do not change their assessments with more time; and they have high agreement with each other on who is attractive (Olson & Marshuetz, 2005; Willis & Todorov, 2006). Members of different cultures show agreement on attractiveness while favoring features unique to their own culture (Cunningham, Roberts, Barbee, Druen, & Wu, 1995; Jones, 1995). Even babies and young children show preferences towards attractive people (Langlois et al., 1987; Langlois et al., 2000; Salvia, Sheare, & Algozzine, 1975). Taken together, these results indicate that the perception of attractiveness has a biological basis in specialized perceptual processing that is automatic and stimulus-driven.

This conclusion is further supported by evidence that human beings have evolved to perceive as attractive potential mates who possess discernable indicators of fertility, hormonal and immune system health, social desirability, and other signs of reproductive fitness (Law Smith et al., 2006; Soler et al., 2003; Thornhill & Gangestad, 1999; Tovée, Hancock, Mahmoodi, Singleton, & Cornelissen, 2002). Indeed, much of the research on attractiveness concerns the biological markers of beauty, sex appeal, and relationship potential and their selective advantage (e.g., Conway, Jones, DeBruine, & Little, 2008; for reviews see Etcoff, 1999; Rhodes & Zebrowitz, 2002).

Youth, vitality, skin quality, bilateral symmetry, familiarity, and nearly ideal facial proportions all enhance attractiveness (Cunningham, 1986; Henss, 1991; Jones, Little, & Perrett, 2004; Langlois & Roggman, 1990; Rhodes & Tremewan, 1996; Rhodes, Proffitt, Grady, & Sumich, 1998). Bilateral symmetry in men, for example, is correlated with running speed, resistance to disease and parasites, sperm quality and count, healthy hormonal levels, and mental well-being (Manning & Pickup, 1998; Manning, Scutt, & Lewis- Jones, 1998; Manning, Gage, Diver, Scutt, & Fraser, 2002; Thornhill & Gangstead, 1993). People have an especially strong preference for bilateral symmetry in human faces (Little & Jones, 2003). Hence, it is possible that we perceive symmetrical faces as attractive, because we inherited perceptual mechanisms favoring symmetry from our ancestors who made reproductively successful mate choices. By extension, the selective pressure to perceive as unattractive those lacking in reproductive fitness may have led to the evolution of the perceptual and cognitive mechanisms responsible for the feelings of aversion associated with the uncanny valley.

1.2.2. Explanations involving cognitive processing

1.2.2.1. The cognitive dissonance of liminal objects: The focus of evolutionary aesthetics research is on identifying aesthetic norms that are universal across cultures and rooted in human biology (Rhodes et al., 2001). However, because *Homo sapiens* did not evolve with robots or animated characters, it is especially important not to overlook the ways in which our evaluative standards are socially constructed within the constraints of human biology. What is potentially most disturbing about artificial human forms is not how they look but what they signify: a challenge to their maker's uniqueness. Robots and CG characters are liminal objects, lying on the boundary of human and nonhuman, calling into question the very distinction (MacDorman et al., 2009; Turkle, Taggart, Kidd, & Daste, 2006; Turkle, 2007). Needless to say, the ensuing dissonance is more cognitive than perceptual and is likely to engage brain regions indentified with motivated as opposed to "cold" reasoning (i.e., the ventromedial prefrontal, anterior cingulate, posterior cingulate, insular, and lateral orbital cortices; Westen, Blagov, Harenski, Kilts, & Hamann, 2006). Thus, cognitive dissonance and certain kinds of uncanny valley experiences may have common neural underpinnings (Pollick, in press).

1.2.2.2. Sorites paradoxes involving personal and human identity: Ramey (2005) argues that the uncanny valley is caused by the linkage of two qualitatively distinct categories– human and robot—by a quantitative metric (i.e., degree of human likeness) that undermines their original separation. The same valley may appear whenever one kind of thing changes "little by little" into a different kind of thing, as in the transformation of an ovum into a human being. Can we determine precisely when human life begins? In the abortion debate, moral uncertainty about what lies between the clearly cellular and clearly human is particularly disturbing, because we identify ourselves with the human end of the continuum. Figures in Mori's graph perceived to lie between robot and human may be disturbing for similar reasons. Moreover, our identification with these robot-machine hybrids may beg the question, "Aren't we all just machines?" This question may excite unconscious fears of annihilation because, if we are just machines, then we are also mortal machines, that is, machines without hope for continuation after death.

1.2.2.3. Terror management theory: Terror management theory has demonstrated how subliminal reminders of death can cause a pervasive shift in our attitudes and preferences (Solomon, Greenberg, & Pyszczynski, 1998). In particular, these reminders cause us to favor those who have opinions that support our cultural worldview. According to Becker (1973), cultural worldviews give our lives meaning and permanence in part by offering a literal or symbolic transcendence of death to those who live up to their standards (Pyszczynski, Greenberg, & Solomon, 1999). Solomon et al. (1998) hypothesize that cultural worldviews reduce the anxiety caused by our uniquely human awareness of our own mortality. But as social constructions, cultural worldviews are fragile compared to our more visceral fear of annihilation and, therefore, are in need of support.

A complementary explanation is that personal identity is socially constructed in terms of our cultural worldview. Identifying the self with something larger and seemingly more permanent–be it family, nation, God, or an immortal soul—provides solace in the face of death. The presence of android robots or CG characters in society, and the mechanistic view of human behavior that they engender, challenges human uniqueness and consequently undermines our sense of personal and human identity (MacDorman et al., 2009). Hence, it is unsurprising that an uncanny android can elicit the same psychological defenses as subliminal reminders of death (MacDorman & Ishiguro, 2006).

1.3. Past investigations of the uncanny valley

There are many ways to conceptualize human likeness. The independent axis of Mori's graph offers one method. A character's form (e.g., shape, texture), dynamics (e.g., motion quality, gestures, facial expressions, speech, intonation, tone of voice), and interactivity (e.g., timing, contingency) could all be varied in their degree of human likeness. Thus, human likeness may be operationalized in terms of formal, quantifiable properties. Individual differences can also influence human perception of CG characters. These include differences that are physiological (e.g., genetic, developmental, sensory acuity, age), cognitive (e.g., learning, habituation, traumatic experiences), social (e.g., childhood experiences, relationships, fetishism), and cultural (e.g., techno-friendly versus technophobic societies) in origin. All of these factors and the relations among them are crucial to understanding the aesthetic dimensions of CG characters. Furthermore, there are many ways to conceptualize *shinwakan*—the dependent axis of Mori's graph. *Shinwakan* roughly translates as a feeling of rapport, and Mori identifies negative *shinwakan* with eeriness (*bukimi*). Human likeness, rapport, and eeriness could also be operationalized in a number of different ways. For example, an increase in skin conductivity as measured by galvanic skin response could potentially indicate an eerie stimulus—as may neural activity in the amygdala as measured by fMRI.

This empirical study is limited to exploring whether an uncanny valley exists for CG characters in still images. It focuses on how facial proportions, skin texture, and levels of detail affect the perceived eeriness, human likeness, and attractiveness of CG characters as indicated on self-reported semantic-differential scales. It also explores how skin textures and levels of detail affect human sensitivity to CG facial proportions as indicated by the degree of interrater agreement. These are important issues. Changing the lower face height or the position of the chin, upper lip, or jaw by as little as 1 mm in a profile can make a human face look unacceptable (Giddon, Sconzo, Kinchen, & Evans, 1996). Animators of human photorealistic CG characters need to know which facial features have low tolerances so that they can take precautions to keep their characters out of the uncanny valley. The results of this study are examined in light of possible explanations of the uncanny valley. Because the results indicate some causes of the uncanny valley, this study proposes a few design principles for bridging it.

Our interest in the uncanny valley arose from the goal of building robots that are indistinguishable from human beings for use in social and cognitive science experiments with people (MacDorman et al., 2005). MacDorman (2006) produced a valley in self-reported ratings on a strange–familiar scale by morphing head-shots of a humanoid robot, an android, and a human. Eeriness ratings were highest in the region of the valley that fell between the robot-looking humanoid and human-looking android. In a follow-up experiment, Hanson (2006) showed the valley could be bridged by carefully designing the steps between robot and android. One flaw of these studies is that the formal properties of the face were not varied systematically along clearly defined dimensions (e.g., polygon count, texture, eye separation) to isolate the elicitors of self-reported eeriness. In addition, the morphing technique introduced visual artifacts that could have increased eeriness ratings.

Our next study examined participants’ evaluations of a diverse range of stimuli, comprising 11 images of people, androids, mechanical-looking robots, and two- and three-dimensional depictions of people and robots (Green, MacDorman, Ho, & Vasudevan, 2008). Participants used a Flash application to modify the proportions of each face along one of four facial dimensions: face height, cheek width, eye separation, and jaw width. Face height (i.e., eye height) and cheek width were selected, because Cunningham (1986) found in a regression analysis of 21 facial proportions of 50 women that these two dimensions—in addition to nose area and smile width—accounted for 50% of the variance in attractiveness ratings. (Nose area and smile width were excluded, because some of the robot stimuli lacked a discernable nose or mouth.) Grammer and Thornhill (1994) found that different facial proportions influenced the perception of attractiveness, dominance, sexiness, and health. Prominent eyes and cheekbones contributed most to males’ evaluations of females, while jaw width and lower-face proportions contributed most to females’ evaluations of males. Based on their results, we added eye separation and jaw width to our list of dimensions along which to vary facial proportions.

In Green et al. (2008), participants indicated a range of values for each facial proportion that appeared acceptable for each face. Against expectations, the acceptable range did not narrow significantly across all four proportions for more humanlike faces. Because the stimuli were derived from 11 very different faces, facial variations other than the faces' degree of human likeness may have prevented a consistent pattern from emerging from the data. For this reason, the present study starts with a single computer graphics based model of a human face and systematically decreases the face's human photorealism by reducing the skin texture photorealism and level of detail (i.e., number of polygons or lines) or by moving facial proportions away from human norms. In Green et al. (2008), participants also indicated which facial proportions looked best by adjusting the face height, eye separation, cheek width, and jaw width of the 11 faces. The results showed heightened participant sensitivity to the best point as human likeness increased, as measured by interrater agreement. The same methodology is applied to the base model in this study for face height and eye separation. (Cheek width and jaw width were excluded from consideration in this study, because a change in skin texture can affect how changes in these dimensions are perceived.)

In an experiment that morphed faces of dolls, masks, and CG characters into human faces, Seyama and Nagayama (2007) failed to uncover evidence of an uncanny valley for nearly human-looking characters. However, a valley became evident when they enlarged the characters' eyes. In particular, the combination of a 50% increase in eye size and human texture and proportions resulted in much greater perceived eeriness than the same proportions with a doll's texture. These results are important to CG animators, because they indicate that exaggerating the size of the eyes, which artists typically do (Costa & Corassa, 2006), may backfire when using a photorealistic facial texture.

The eeriness of oversized eyes combined with a photorealistic skin texture may indicate that eeriness is not the result of a certain degree of human likeness, but the result of a discrepancy between more human-looking and less human-looking elements. This mismatch hypothesis can be traced to Mori's original article in which he notes the eeriness of a prosthetic hand that looks natural but feels artificial. The hypothesis is supported by our study of robot videos, which found that robots that possessed both human and nonhuman characteristics elicited fear, disgust, anxiety, dislike, and shock–the emotions associated with eeriness (Ho, MacDorman, & Pramono, 2008). It is also supported by a study, which found that a certain motion performed by more human-looking characters is seen as *less* natural than the same motion performed by animated CG characters. This negative response bias shows a significant correlation with brain activity as detected by fMRI in shared circuits for mentalizing, including the left temporoparietal junction and the anterior cingulate cortex (Chaminade et al., 2007). Studies on people's feelings of copresence when interacting with a character in a virtual environment also show that copresence is lowest when there is a mismatch between the character's perceived human photorealism and perceived human behavioral fidelity (Bailenson et al., 2005; Nowak & Biocca, 2003). Vinayagamoorthy, Steed, and Slater (2005) note that virtual characters will be evaluated more positively when their degree of human behavioral fidelity is consistent with their degree of human photorealism.

This paper presents four empirical studies that explore issues related to the uncanny valley in still images:

I. *Baseline human likeness and eeriness*: Participants rated the eeriness and human likeness of a 3D model of a male human head presented at three different textures and levels of detail. The motivation for the survey was to determine whether rendering a CG face in more detail would cause it to look eerier. The survey also provided baseline human likeness ratings for each texture and level of detail to be used in the next study.

II. *Sensitivity to best proportions*: Participants independently manipulated the eye separation and face height of the 3D model to determine which proportions looked best. The change from the original proportions were plotted against the human likeness ratings of the baseline survey to determine whether, as perceived human likeness increased, the proportions that looked best converged on the human model's original proportions.

III. *Eeriest level of detail*: For extreme facial proportions, participants selected the level of detail at which the CG face looked eeriest. Our expectation was that people would be more disturbed by extreme facial proportions on more detailed CG faces, because detailed faces more strongly enlist specialized human facial processing, which would apply more stringent human aesthetic norms to the evaluation of the faces.

IV. *Eyes–face mismatch*: Participants rated the eeriness, naturalness, and attractiveness of CG faces at five levels of eye photorealism and five levels of skin photorealism with normal-sized eyes and eyes enlarged by 50%. The idea was to determine whether a mismatch in the level of eye and skin photorealism increased eeriness and whether eye enlargement increased the eeriness of more photorealistic faces by a greater extent than less photorealistic faces.

2. Study I: Baseline eeriness and human likeness

This study collected eeriness and human likeness ratings of photorealistic, bronze, and line drawing renders of the base model at 11 levels of detail. More detailed renders of more photorealistic textures were predicted to look more human. One purpose of this study was to establish baseline eeriness and human likeness ratings of the stimuli used in Studies II and III. In addition, the study was intended to check the prediction of the uncanny valley graph that eeriness would reach a peak near total human likeness (Mori, 1970). Human likeness was gauged both objectively, by texture photorealism and level of detail, and subjectively, by participant ratings of human likeness.

2.1. Hypotheses

One aim of this study was to confirm whether the stimuli would function as expected. Because the original model was designed to look as photorealistic as possible, and all stimuli —including the stimuli with less detail or skin photorealism—were derived from this model, ratings of human likeness were expected to increase with the level of detail and texture photorealism.

H1A. Human texture: CG faces rendered with a more photorealistic texture are perceived as more human than those rendered with a less photorealistic texture.

H1B. Human detail: CG faces rendered in more detail are perceived as more human than those rendered in less detail.

As potential explanations of the uncanny valley, theories concerning the role of disgust in pathogen avoidance or facial and body proportions as indicators of fertility in mate selection draw on evidence of specialized perceptual mechanisms for evaluating human faces and bodies (MacDorman & Ishiguro, 2006). Presumably, these human-specific mechanisms would provide more consistent evaluations of stimuli than more general mechanisms for evaluating objects. Therefore, sensitivity to the human likeness of a CG face is expected to increase with the perceived human likeness of the face. In this study, sensitivity is operationalized as interrater agreement.

H2. Sensitivity human: Sensitivity to human likeness increases with perceived human likeness.

Mori's uncanny valley graph predicts forms that are close to human appearance will be most eerie. Therefore, perceived eeriness is predicted to be high when human likeness is also high. This study operationalizes human likeness as level of detail, texture photorealism, and self-reported ratings of human likeness.

H3A. Eerie texture: CG faces rendered with more texture photorealism are eerier than those rendered with less texture photorealism.

H3B. Eerie detail: CG faces rendered in more detail are eerier than those rendered in less detail.

H3C. Eerie human: CG faces perceived as more humanlike are eerier than those perceived as less humanlike.

2.2. Methods

2.2.1. Participants—Participants for all studies were recruited by e-mail using a random sample from a list of 126,425 undergraduate students and recent graduates from eight campuses administered by a Midwestern university. In total, there were 3294 participants in all of the studies.

In the *baseline–eerie* survey, there were 458 participants: 81.2% were 18–25, 60.0% were female, and 95.9% were US born. The confidence level was 95% with a ±4.57% error range. In the *baseline–humanlike* survey, there were 407 participants: 78.4% were 18–25 years old, 63.9% were female, and 92.9% were US born. The confidence level was 95% with a ±4.85% error range.

2.2.2. Stimuli—Our previous studies on the effects of varying facial proportions used dissimilar base figures (Green et al., 2008). To improve the experimental control of the stimuli, a photorealistic 3D model of a male human head was developed. From this photorealistic model two additional models with different skin textures were derived:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

metallic bronze with simplified eyes and a line drawing. The top row of Fig. 2 shows the models with these three skin textures.

The degree of photorealism in each model was varied by changing the level of detail. To decrease the level of detail for the photorealistic and bronze models, the number of polygons was reduced and smoothing was removed. For the line model, the number of lines was reduced. All participants in Study I viewed a total of 33 stimuli: the line, bronze, and photorealistic texture models at 11 levels of detail each. For each texture the top row of Fig. 2 shows the model at the highest level of detail (level 11), the middle row shows the model at the median level of detail (level 6), and the bottom row shows the figure at the lowest level of detail (level 1). (Levels 2–5 and 7–10 are not pictured.)

2.2.3. Procedures—To establish a baseline of the effects of the level of detail, a pilot study was conducted. Participants were directed to a website that presented images with differing textures and levels of detail with facial features in their original proportions. Participants rated each image on an 11-point semantic differential scale for either eeriness (458 participants) or human likeness (407 participants). The presentation order of the 33 stimuli was randomized for each participant. The eeriness scale was anchored at *totally reassuring* (−5) and *totally eerie* (+5). The human likeness scale was anchored at *totally nonhuman* (−5) and *totally human* (+5).

2.3. Results and discussion

Fig. 3 shows that on average the photorealistic texture models were rated as more humanlike than the bronze texture models, which were in turn rated as more humanlike than the line textured models. The error bars indicate the 95% confidence interval (CI). A one-way analysis of variance (ANOVA) confirmed that the difference between the photorealistic texture and the bronze texture was highly significant at all 11 levels of detail ($F(1, 814)$ ranged from 168.12 to 593.33, $p = .000$) as was the difference between the photorealistic texture and the line texture ($F(1, 814)$ ranged from 206.16 to 1138.93, $p = .000$). In support of H1A, these results indicate that CG faces rendered with a more photorealistic texture are perceived as more human. The increase in human likeness was much greater when using a photorealistic texture in place of a bronze texture than when using a bronze texture in place of a line texture. A one-way ANOVA confirmed that the difference between the bronze texture and line texture only reached significance at level 2 and 8–11 with $F(1, 814)$ ranging from 6.60 to 58.20 and p ranging from .010 to .000.

In support of H1B, ratings of human likeness increased as the level of detail increased for the bronze and photorealistic texture CG faces. The correlation was highly significant for the bronze ($r = .16$, $p = .000$, two-tailed) and photorealistic texture models ($r = .45$, $p = .000$, two-tailed). The effect size was small for the bronze texture model but large for the photorealistic texture model.

For the line texture, ratings of human likeness increased as the level of detail increased only to the midpoint (level 6). From the midpoint, human likeness decreased as the level of detail increased, which is counter to H1B. After the midpoint, adding more lines made the line texture face look less humanlike. In addition, the pupil and iris do not appear at the first two levels of detail in the line texture model. This could have a large effect given the importance of eyes in social communication (Emery, 2000). Thus, it is unclear whether the low human likeness ratings for the first two levels should be attributed to the model's overall level of detail or the fact that the eyes were missing. The correlation between the level of detail and the ratings of human likeness was approaching significance for the line texture model (r = .03, p = .051, two-tailed); however, the effect size was almost negligible.

Interrater agreement (r_{wg}) (James, Demaree, & Wolf, 1993) on human likeness for the bronze, line, and photorealistic textures was.55,.55, and.95, respectively. Clearly, participant agreement on human likeness was much higher for the photorealistic texture. Another indication of interrater agreement is the association of the mean and standard deviation of human likeness ratings. Fig. 4 shows increased ratings of human likeness are associated with decreased standard deviations for the photorealistic texture, but not for the line and bronze textures. The correlation between the level of detail and the standard deviation in ratings of human likeness was not significant for the line texture model (r = − .02, p =.664); however, it was significant for the bronze texture model (r = −.11, p = .032) and highly significant for the photorealistic texture model (r = −.56, p = .000). All three correlations were negative, and the effect size was large for the photorealistic texture model, but small for the bronze texture model and almost negligible for the line texture model.

Taken together, these results support H2 for the photorealistic texture model only: sensitivity to human likeness increases with perceived human likeness only for this model. These results indicate that only the photorealistic texture elicits the specialized perceptual processing associated with increased agreement on human likeness.

Fig. 5 shows mean ratings of eeriness were lowest for the photorealistic texture. These results are contrary to H3A, which had predicted–in accordance with Mori's (1970) uncanny valley graph–that CG faces with higher texture photorealism would be perceived as more eerie. A one-way ANOVA confirmed that the difference in eeriness ratings between the photorealistic texture and bronze texture was highly significant at all levels (for level 1 F(1, 914) = 11.95, p = .001 and for levels 2 through 11 F(1, 914) ranged from 14.93 to 64.20, p = .000). Except for level 3, the difference between the photorealistic texture and the line texture was highly significant at all levels (F(1, 914) = 9.73, p = .002 at level 4 and F(1, 914) ranged from 19.45 to 376.30, p = .000 at the remaining levels). The difference between the bronze texture and line texture was highly significant at level 3 (F(1, 914) = 10.30, p = .001) and levels 8 through 11 (F(1, 914) ranged from 30.12 to 191.77, p = .000). It was significant at level 1 (F(1, 914) = 5.89, p = .015) and failed to reach significance at the remaining levels.

Fig. 5 also shows mean ratings of eeriness decreased as the level of detail increased for the bronze and photorealistic textures. This is contrary to H3B, which had predicted—in accordance with the uncanny valley graph—that CG faces rendered in more detail are eerier. For the line texture, ratings of eeriness decreased initially, then increased as the level of detail increased. The initial decrease, however, could be attributed to the fact that the line model had no eyes at the first two levels of detail, which could have contributed to their higher eeriness ratings. The correlation between the level of detail and the ratings of eeriness was highly significant for the line (r = .11, p = .000), bronze (r = −.19, p = .000), and photorealistic texture model (r = −.26, p = .000). The correlation was positive for the line texture model but negative for the bronze and photorealistic texture model.

Fig. 6 depicts an inverse relation between mean eeriness ratings and mean human likeness ratings. This trend is contrary to H3C, which predicts—in accordance with the uncanny valley graph—higher perceived eeriness at higher perceived human likeness. Increased level of detail led to ratings of increased human likeness and decreased eeriness for the bronze and photorealistic textures (Fig. 6). The ratings for the line textured figure almost doubled back on themselves. The correlation between the ratings of human likeness and eeriness cannot be calculated, because they involve different sets of participants. However, the correlation between the average ratings of human likeness and the average ratings of eeriness was highly significant for the line (r = −.76, p = .006), bronze (r = −.99, p = .000), and photorealistic texture model (r = −.99, p = .000). All three correlations were negative, and their effect sizes were large.

An important question concerns why CG faces with higher texture photorealism, increased detail, and higher human likeness ratings tended to be less eerie, when the uncanny valley graph seems to predict they would in fact be more eerie. One interpretation is that the graph's prediction is simply is false. However, there is another, more likely possibility. The detailed, photorealistic CG face is designed to resemble a human face, and a human face is in a sense the product of evolutionary design—co-evolution involving selective processes, such as mate selection and reproductive fitness, that affect both human perceptual mechanisms and facial morphology. The various examples Mori (1970) provides of points along his graph are the product of artistic design—humanoid robots and *bunraku* puppets, for example. But in this study applying a bronze or line texture to the model introduced visual artifacts that were not the product of an evolutionary or artistic design process. In decreasing the polygon count in the less detailed CG faces, for example, the placement of edges between polygons did not involve the judgment of a trained artist. All images were created in Maya, and the level of detail was adjusted using automated tool settings. Clearly, tool settings are not a substitute for the trained eye of an artist. This complicates the interpretation of the results.

The outcome could be quite different, if we had started with, say, a cute, cartoony Barbie doll and had added a photorealistic human texture. When adjusted to human scale, Barbie's waist is sixtenths that of a woman with anorexia (Norton, Olds, Olive, & Dank, 1996). If a photorealistic texture elicits strong expectations of typical human proportions, a photorealistic Barbie doll could look very eerie. The Study I results indicate the photorealistic texture had the largest effect of any treatment. Faces with a photorealistic texture were rated as more humanlike and less eerie—with greater interrater agreement on human likeness—than their bronze and line counterparts. The photorealistic texture may be engaging specialized perceptual processing to a greater extent than the bronze and line textures. This processing could result in an averse reaction to an out-of-proportion face with a photorealistic texture. The relation between extreme proportions and the uncanny valley is explored in Studies III and IV.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3. Study II: Sensitivity to best proportion

Hanson (2006) and MacDorman and Ishiguro (2006) argue that acceptable norms of attractiveness narrow as a form comes to look human. A face that looks substantially human but with nonhuman imperfections may be human enough to elicit the perceivers' innate or acquired model of a human other without being able to satisfy the model's expectancies in certain respects. If the range of what is acceptable narrows, because the stimulus has elicited specialized face processing, it seems reasonable to assume that agreement on what is most attractive should increase. Green et al. (2008) found sensitivity to facial proportions increased with the human likeness of the face. But that study varied the human likeness of the faces by using different faces, which introduces extraneous variation. It is important to reproduce the results of that study by varying systematically the human likeness of the same face.

An experimental apparatus was devised in the form of a Flash application accessible from a website. Participants were able to adjust a face along a facial dimension to determine and select which proportions looked best. As in Study I, participants applied this procedure to CG faces that varied in texture photorealism and level of detail.

3.1. Hypotheses

> *H4. Human original*:As human likeness increases, the best-looking facial proportions lie closer to the original proportions of the human model.

The basic idea is that if a face looks human, it should elicit perceptual processing that is specialized for faces. Given that the attractiveness of a face can elicit an affective response that increases reproductive fitness, the range of what looks acceptable is expected to be narrower for faces than for other kinds of objects. Therefore, if a human model is chosen that already conforms to human norms, people would be likely to set facial proportions close to their original values. However, if the face is made to look less human—for example, by using a bronze or line texture instead of a photorealistic one or by using fewer polygons—people might set the proportions to be further from their original values. Without the strong activation of a human model, proportions that look ideal could drift from human norms with perhaps other aesthetic principles coming into play instead. We see this in cartoon depictions of human beings, which can be beautiful despite having grossly diminutive or exaggerated proportions (e.g., Jessica Rabbit in the film *Who Framed Roger Rabbit*).

> *H5. Sensitivity proportion*: The more human CG faces look, the greater the sensitivity to facial proportions, as measured by interrater agreement on what proportion looks best.

The degree to which different people agree on what proportions look best is another way of approaching the issue of whether more stringent norms are applied to forms that look more human. If a more human-looking face is able to enlist more specialized processing, it is reasonable to expect greater consistency in what proportions different people judge to look best. For method triangulation, both hypotheses are explored.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3.2. Methods

3.2.1. Participants—In the *sensitivity–proportions* experiment, there were 1118 participants: 74.4% were 18–25 years old, 60.1% were female, and 89.9% were US born. The confidence level was 99% with a ±3.84% error range.

3.2.2. Stimuli—The experimental apparatus sequentially presented participants with 18 faces that varied along three dimensions. Each face varied by texture photorealism (photorealistic, bronze, and line drawing) and level of detail (low, medium, and high). In addition, the participant could manually adjust the face along a third dimension: either eye separation or face height.

Each of the 18 adjustable CG faces was implemented as a Flash video embedded in a Flash application. These Flash videos were never played as such. Rather, the video format was the internal representation used for the stimuli. When participants manually varied a model's facial proportions, at the implementation level, they were stepping backwards or forwards through the frames of these videos. Each video modified one facial proportion, eye separation or face height, from −10% to +10% of the original value. Each video contained 41 frames with each frame representing a 0.5% change in proportion. Thus, there were 738 stimuli in total (18 videos × 41 frames).

Fig. 7 depicts the points used for determining the facial dimensions. Eye separation is defined as the distance between the pupils divided by the width of the face at the cheekbones:

$$\frac{E2 - E1}{C2 - C1} \quad (1)$$

Face height is defined as the length of the lower face divided by the height of the head. The length of the lower face is the vertical distance from the midpoint of a line between the pupils to the chin; head height is the distance from the top of the head to the chin. Thus the formula for face height is

$$\frac{F3 - F2}{F3 - F1} \quad (2)$$

3.2.3. Procedures—Eighteen faces were sequentially presented, representing three textures, three levels of detail, and two facial dimensions. Each faces could be manually adjusted in either its eye separation or face height. The presentation order and model's initial facial proportion were randomized for each participant. Participants were instructed to use the left and right arrows to adjust the facial proportion until it looked best and to indicate the best proportion by pressing a button. (They could either click on the arrows using their mouse pointer or press the arrow keys on their keyboard.) All 1118 participants could have potentially viewed all 738 stimuli (41 frames × 18 videos) by manipulating the arrows.

3.3. Results and discussion

A one-way ANOVA shows a highly significant difference between the points selected as best both by facial texture and level of detail ($F(2, 2646) = 102.41$–183.40, $p = .000$). The mean best points approach the original facial proportions with increased human likeness (Fig. 8). The *x*-axis of Fig. 8 is the humanlike rating from the baseline survey.

The models are ranked from lowest to highest human likeness as follows: line–low detail, line–high detail, bronze–low detail, line–middle detail, bronze–middle detail, bronze–high detail, photorealistic–low detail, photorealistic–middle detail, and photorealistic–high detail. The mean best points varied widely for the first few stimuli, which were rated as more nonhuman than human. After that, the best point for eye separation gradually decreased, approaching the actual proportion for the faces rated as nearly human. A similar pattern occurred for face height, though participants wanted the most humanlike image to have a slightly taller face, perhaps because the hairline was placed too high in the CG model. Apart from this, the photorealistic baseline CG models appeared to be close to the human ideal for face height and eye separation. These results support H4: with increasing human likeness, the facial proportions rated as *best* drew closer to the human model's original facial proportions.

The more participants agree on which facial proportion looks best, the lower the standard deviation becomes in the proportions they determine are best. The standard deviation of the best facial proportion thus provides a measure of human sensitivity to facial proportions (interrater agreement); however, sensitivity only increased slightly with human likeness, and an analysis of variance of absolute *z*-scores found no significant relation between human likeness and sensitivity to facial proportions.

For facial texture a significant difference exists in the relation between sensitivity and face height ($F(2, 10057) = -3.99$, $p < .05$, $\omega = 0.02$). Tanhame post hoc tests indicate more sensitivity to the bronze ($MD = 0.08$, $SE = 0.009$, $p = .000$) and photorealistic textures ($MD = 0.10$, $SE = 0.010$, $p = .000$) than to the line texture. H5 predicted that as human likeness increased, participants would be more sensitive to what facial proportions looked best, as measured by interrater agreement. H5 was supported for texture photorealism but not level of detail.

Green et al. (2008) found sensitivity to facial proportions was related to the degree of human likeness of the figure. But that study used different kinds of stimuli (photographs of real, 3D CG, and 2D humans and robots). In this study texture proved to be more important than the level of detail in determining sensitivity to the best facial proportion. On the most photorealistic face, participants set the eye separation very near to the model's actual proportion. Texture could be crucial in enlisting facial processing in the brain that applies human norms to facial proportions.

4. Study III: Eeriest level of detail

Ho et al. (2008) observed that a robot is eeriest when human elements create an expectation of a human form that nonhuman elements fail to satisfy. Therefore, it is worthwhile to determine whether extreme facial proportions, which are far from human norms, would be perceived as eerier at higher levels of detail. This would support evolutionary explanations–based on mate selection and pathogen avoidance–that evolved or acquired perceptual mechanisms endow human beings with heightened sensitivity to defects in human-looking faces (MacDorman & Ishiguro, 2006).

4.1. Hypothesis

H6A. Eerie detail, eye separation off: Setting the eye separation of a CG face to ±10% will cause the face to be perceived as eeriest at a higher level of detail than at its original eye separation.

H6B. Eerie detail, face height off: Setting the face height of a CG face to either ±10% will cause the face to be perceived as eeriest at a higher level of detail than at its original face height.

4.2. Methods

4.2.1. Participants—In the *sensitivity–eeriness* experiment, there were 852 participants: 75.5% were 18–25, 62.0% were female, and 89.1% were US born. The confidence level was 99% with a ±4.40% error range.

4.2.2. Stimuli—The experimental apparatus sequentially presented participants with 15 CG faces that varied along three dimensions. Each face varied by texture photorealism (photorealistic, bronze, and line drawing) and facial proportion (original, widely-set eyes, narrowly-set eyes, low face height, and high face height). In addition, the participant could manually adjust the face along a third dimension: level of detail. In selecting the eeriest level of detail, the participant had 11 levels to choose from (as in Studies I and II), ranging from low to high.

The 15 CG faces were implemented as video data embedded in a Flash application at a website. Five videos were created for each of the three textures, varying the level of detail from low to high. A video was created for the original face and two extremes (±10%) along the two facial dimensions: eye separation and face height. Fig. 9 shows these four extremes for the CG model with a photorealistic texture and high level of detail. Each video contained 11 frames with each frame representing a 10% step between low and high levels of detail.

4.2.3. Procedures—Fifteen CG faces were sequentially presented, representing three textures at five facial proportions. The presentation order was randomized for each participant, and the face's initial level of detail was randomized for each face. Participants were instructed to use the left and right arrows to adjust the level of detail until it looked eeriest.

4.3. Results and discussion

Fig. 10 shows the level of detail at which each texture and facial proportion was deemed eeriest, ranging from lowest (1) to highest (11). The error bars indicate the 95% confidence interval (CI). The frame rated as eeriest for the normal facial proportion was lower (less humanlike) than for the extreme (±10%) facial proportions.

A one-way ANOVA for the eeriest point indicates a highly significant difference between the frame selected as eeriest for all three facial textures ($F(2, 2646) = 102.41–183.40$, $p = .000$). Sensitivity to eeriness reached significance for the bronze and photorealistic facial textures ($F(2, 2646) = 5.51–25.47$, $p = .000–.004$). Results of the ANOVA are detailed in Table 1. Post hoc tests give a possible explanation for the low effect sizes. Tamhade's T2 shows the difference between the normal position and each extreme is significant (all *p*s = .000), while the differences between extremes were not significant. These results support H6A and H6B: eye separation or face height that lies ±10% from the original causes the CG face to appear eeriest at a higher level of detail.

A two-way ANOVA was used to determine whether there was an interaction effect between the level of detail and the texture of the model (line, bronze, or photorealistic). The results show that there was a highly significant interaction effect for eye separation ($F(4, 7100) = 80.88$, $p = .000$) and for face height ($F(4, 7100) = 11.20$, $p = .000$). For eye separation there was a highly significant single effect for the model's texture ($F(2, 7100) = 130.33$, $p = .000$) and for the level of detail ($F(2, 7100) = 88.13$, $p = .023$), considered separately. For face height there was also a highly significant single effect for the model's texture ($F(2, 7100) = 47.76$, $p = .000$) and for the level of detail ($F(2, 7100) = 20.45$, $p = .000$).

In sum, faces with normal proportions were rated eeriest at lower levels of detail. Faces with extreme proportions (±10%) were rated eeriest at higher levels of detail–that is, when the face was more humanlike. These results hold for all facial textures and clearly support H6A and H6B.

The only significant differences in sensitivity were for the photorealistic texture between original and narrowly-set eyes ($MD = −0.08$, $SE = 0.028$, $p < .05$), and original and low face height ($MD = − .11$, $SE = 0.027$, $p = .000$). This indicates greater sensitivity to the normal face. Sensitivity to eeriness in the line texture did not follow the photorealistic and bronze textures. This is probably because low and high levels of detail were both perceived as eerier and less humanlike than the midpoint (Figs. 3 and 5).

5. Study IV: Eyes–face mismatch

This study further explores how a mismatch in the human likeness of different elements of a character can cause it to seem eerie (Ho et al., 2008). The photorealism of the CG model's eyes and skin are varied independently to determine whether a matching level of photorealism between eyes and face can reduce eeriness and increase naturalness and attractiveness. This study also explores the influence of enlarged eyes on eeriness, naturalness, and attractiveness when eye and skin photorealism are varied independently. Seyama and Nagayama (2007) reported that, for photorealistic facial textures, enlarged eyes greatly increased eeriness.

5.1. Hypotheses

H7. Eerie texture mismatch: A CG face will look less eerie when the texture of the eyes and skin are at a similar level of photorealism than when their level of photorealism differs greatly.

H8. Eerie big eyes with photorealism: The more photorealistic a CG face is the more a 50% enlargement of the eyes will increase its perceived eeriness, artificiality, and ugliness.

5.2. Methods

5.2.1. Participants—In the first half of this study, there were 302 participants: 75.5% were 18–25, 72.8% were female, and 95.0% were US born. The confidence level was 95% with a ±5.63% error range. In the second half of this study, there were 157 participants: 68.3% were 18–25, 69.7% were female, and 92.9% were US born. The confidence level was 95% with a ±7.82% error range. Different participants were used in the first and second half of this study to avoid habituation and fatigue effects.

5.2.2. Stimuli—To create the 50 faces used in the first half of this study, the detailed bronze model with simplified blue eyes was overlayed with the detailed photorealistic model (Fig. 2, top left and top center). Next the opacity of the detailed photorealistic model was varied for the skin and eyes independently. It was varied from 0% to 100% at 25% increments. This resulted in five levels of photorealism for the eyes and five levels of photorealism for the skin for a total of 25 different combinations. The eyes of the detailed photorealistic and bronze models were enlarged 50%, and the above process was repeated to create an additional 25 faces.

The same procedure was used to create the 50 faces used in the second half of this study. The only difference was that the detailed photorealistic and bronze base models' appearance was enhanced: The photorealistic model was given photorealistic hair; the bronze model was given cartoon hair; the ears of both models were reduced in size to match their human counterpart; and the eye color of the bronze model was changed to brown (Fig. 12A, bottom left figure).

5.2.3. Procedures—Participants rated 50 faces along three dimensions: eeriness, naturalness, and attractiveness. The presentation order of the 50 faces was randomized for each participant. Participants rated each image on an 7-point semantic differential scale. The eeriness scale was anchored at *very eerie* (−3) and *very reassuring* (+3). The naturalness scale was anchored at *very artificial* (−3) and *very natural* (+3). The attractiveness scale was anchored at *very ugly* (−3) and *very attractive* (+3).

5.3. Results and discussion

In the first half of the study, the CG face with normal-sized eyes was most eerie at 100% eye photorealism and 0% skin photorealism. (One hundred percent photorealism means the texture blended 100% of the photorealistic texture and 0% of the bronze texture.) Its mean rating was −1.52 on the eerie–reassuring scale (SD = 1.24), which is about midway between *slightly eerie* and *moderately eerie*. It was also rated as most artificial (M = −1.97, SD = 1.32) and ugliest (M = −1.23, SD = 1.26).

The CG face was most reassuring at 100% eye photorealism and 75% skin photorealism. Its mean rating was 1.37 (SD = 1.23), which is between *slightly reassuring* and *moderately* *reassuring*. The same face was also rated as most natural (M = 1.93, SD = 1.41). However, the CG face at 75% eye photorealism and 75% skin photorealism was rated as most attractive.

The surface plot of the results for eeriness is convex (Fig. 11A). Similar levels of eye photorealism and face photorealism are rated more positively than dissimilar levels. The surface plots for naturalness and attractiveness were similar to this plot.

Enlarging the eyes by 50% resulted in universally negative ratings on all three dimensions (for eeriness see Fig. 11B). Mean eeriness ratings ranged from −1.60 (SD = 1.22) at 75% eye photorealism and 75% skin photorealism to −2.33 (SD = 0.93) at 100% eye photorealism and 0% skin photorealism. Mean naturalness ratings ranged from −1.37 (SD = 1.56) at 100% eye photorealism and 100% skin photorealism to −2.62 (SD = 0.81) at 75% eye photorealism and 0% skin photorealism. Mean attractiveness ratings ranged from −1.39 (SD = 1.29) at 75% eye photorealism and 50% skin photorealism to −2.14 (SD = 1.05) at 100% eye photorealism and 0% skin photorealism.

Enlarging the eyes by 50% increased mean eeriness ratings most at higher levels of eye photorealism and skin photorealism. Eeriness increased by 2.99 at 100% eye photorealism and 75% skin photorealism, which was the maximum increase. Eeriness increased by 0.82 at 100% eye photorealism and 0% skin photorealism, which was the minimum increase.

In the second half of the study, the CG face with normal-sized eyes was most eerie at 100% eye photorealism and 0% skin photorealism. Its mean rating was −1.62 (SD = 1.36). It was also rated as most artificial (M = −2.08, SD = 1.23) and ugliest (M = −1.17, SD = 1.42).

The CG face was most reassuring at 50% eye photorealism and 75% skin photorealism. Its mean rating was 1.04 (SD = 1.20). The CG face was most natural at 75% eye photorealism and 75% skin photorealism (M = 1.62, SD = 1.36). The CG face was most attractive at 0% eye photorealism and 75% skin photorealism (M = 1.09, SD = 1.11).

The heatmap of the results for eeriness is convex (Fig. 12A). The heatmaps for naturalness and attractiveness were also convex.

Once again, enlarging the eyes by 50% resulted in universally negative ratings on all three dimensions (for eeriness see Fig. 12B). Mean eeriness ratings ranged from −1.54 (SD = 1.13) at 50% eye photorealism and 75% skin photorealism to −2.54 (SD = 0.78) at 100% eye photorealism and 0% skin photorealism. Mean naturalness ratings ranged from −1.58 (SD = 1.29) at 100% eye photorealism and 100% skin photorealism to −2.70 (SD = 0.61) at 100% eye photorealism and 0% skin photorealism. Mean attractiveness ratings ranged from −1.32 (SD = 1.16) at 50% eye photorealism and 75% skin photorealism to −2.25 (SD = 1.01) at 100% eye photorealism and 0% skin photorealism.

Enlarging the eyes by 50% increased mean eeriness ratings between 0.91 and 1.37 at 0% skin photorealism, between 1.38 and 1.85 at 25%, between 1.97 and 2.63 at 50%, between 2.55 and 2.78 at 75%, and between 2.32 and 2.66 at 100%. A similar trend appeared for the other two dimensions: enlarging the eyes 50% increased eeriness, decreased naturalness, and decreased attractiveness, and this effect is much larger for more photorealistic skin textures.

A two-way ANOVA was used to determine whether there was an interaction effect between the level of eye photorealism and skin photorealism. The results show that, for eeriness ratings, there was a highly significant interaction effect between eye photorealism and skin photorealism both for normal eyes ($F(16, 7525) = 27.00$, $p = .000$) and for eyes enlarged 50% ($F(16, 7525) = 6.36$, $p = .000$). For normal eyes there was a highly significant single effect for eye photorealism ($F(4, 7525) = 420.59$, $p = .000$) and for skin photorealism ($F(4, 7525) = 194.94$, $p = .000$). For eyes enlarged 50%, there was also a highly significant single effect for eye photorealism ($F(4, 7525) = 26.94$, $p = .000$) and for skin photorealism ($F(4, 7525) = 21.48$, $p = .000$).

For attractiveness ratings there was a highly significant interaction effect between eye photorealism and skin photorealism both for normal eyes ($F(16, 7525) = 11.81$, $p = .000$) and for eyes enlarged 50% ($F(16, 7525) = 4.10$, $p = .000$). For normal eyes there was a highly significant single effect for eye photorealism ($F(4, 7525) = 366.11$, $p = .000$) and for skin photorealism ($F(4, 7525) = 56.87$, $p = .000$). For eyes enlarged 50%, there was also a highly significant single effect for eye photorealism ($F(4, 7525) = 18.52$, $p = .000$) and for skin photorealism ($F(4, 7525) = 9.44$, $p = .000$).

For naturalness ratings there was a highly significant interaction effect for eye photorealism and skin photorealism both for normal eyes ($F(16, 7525) = 25.77$, $p = .000$) and for eyes enlarged 50% ($F(16, 7525) = 5.19$, $p = .000$). There was a highly significant single effect for eye photorealism ($F(4, 7525) = 995.62$, $p = .000$) and for skin photorealism ($F(4, 7525) = 192.20$, $p = .000$). For eyes enlarged 50%, there was also a highly significant single effect for eye photorealism ($F(4, 7525) = 142.39$, $p = .000$) and for skin photorealism ($F(4, 7525) = 11.18$, $p = .000$).

In the first half of this study, similar levels of eye photorealism and face photorealism were rated more positively than dissimilar levels, which supports H7. This relation appeared as a *synergy ridge* in surface plots of perceived eeriness, naturalness, and attractiveness. The idea of a synergy ridge was first proposed for matching levels of human likeness in appearance and behavior (MacDorman et al., 2005).

A caveat to this study can be explained as follows. The photorealistic model was created to look as similar to the human model as possible. In a sense, the human model reflects evolutionary design as well as cultural attitudes toward appearance (e.g., those concerning corpulence, orthodontics, grooming, hygiene, ornamentation, tribal scarring, body piercing, or lack of these). However, the detailed bronze model was not designed to look like anything. It simply reflects the application of a bronze shader to the photorealistic model in place of human texture maps. Therefore, the appearance of both the bronze and photorealistic model was enhanced for the second half of the study.

Although dissimilar levels of eye–face photorealism are still rated negatively for all dimensions, the results show that 75% skin photorealism is perceived to be least eerie, most natural, and most attractive. More surprisingly, 50% eye photorealism was most reassuring (least eerie) and 0% eye photorealism was most attractive. This shows that backing away from photorealism can sometimes make a CG character less eerie and more attractive.

Enlarging the eyes 50% increased eeriness and decreased naturalness and attractiveness most for more photorealistic skin textures. This result supports H8 and agrees with the findings of Seyama and Nagayama (2007). However, to confirm this effect the experiment should be repeated with faces that vary only in realism, not eeriness, before the eyes are enlarged.

6. Main findings

More human-looking CG faces generally had more photorealistic textures and more detail. When a photorealistic or bronze texture was used, increasing the number of polygons increased human likeness (Fig. 3). However, the graph of human likeness reached a maximum in the line model at the median number of lines. A photorealistic texture caused CG faces to look much more human than a bronze or line textured faces.

Sensitivity to human likeness was heightened and increased with greater human likeness only when a photorealistic texture was used. The more human the photorealistically-textured CG face looked, the easier it was for people to agree on its degree of human likeness (Fig. 4). For a line or bronze texture, sensitivity to human likeness did not increase with human likeness.

Under certain circumstances, less humanlike CG faces can be eerier than more humanlike CG faces. The photorealistic texture was less eerie than the bronze or line texture for nearly ideal facial proportions (Fig. 5). However, a somewhat less than photorealistic texture (75% photorealistic, 25% bronze) was found to be less eerie than the most photorealistic texture in Study IV (Fig. 12A). For nearly ideal facial proportions, increasing the polygon count decreased eeriness. The CG face was less eerie with a photorealistic texture than with a bronze or line texture (Fig. 6). However, the last study showed that a three-to-one blend of the photorealistic and bronze textures was the least eerie (Fig. 12A). Line-textured, bronze, and photorealistic CG faces that looked more human also looked less eerie. However, the second half of the last study showed that the most natural CG face was not the least eerie. The least eerie face had 75% skin photorealism.

As human likeness increased, the best-looking facial proportions were generally closer to the original proportions of the human model (Fig. 8). But contrary to expectations, as the CG character’s level of detail increased, sensitivity to facial proportions, as measured by interrater agreement, did not increase significantly. However, for face height, the bronze and photorealistic textures resulted in significantly higher sensitivity than the line texture.

Facial proportions that are far from ideal look eerier at higher levels of detail than facial proportions that are nearly ideal. Specifically, for a ±10% change in face height or eye separation, the level of detail found eeriest increased. However, contrary to the uncanny valley graph, the eeriest level of detail was still relatively low. This may be an artifact of the stimuli, because the line and bronze models were not designed by an artist, so they look less attractive than a typical hand-drawn model.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

The CG face looked less eerie when the texture of the eyes and skin were at a similar level of photorealism than when their level of photorealism differed greatly (Figs. 11 and 12A). Decreasing photorealism somewhat made the face look less eerie and more attractive (Fig. 12A). Enlarging the eyes by 50% increased the eeriness and decreased the naturalness and attractiveness of a CG face (Figs. 11 and 12B). The increase in eeriness and decrease in naturalness and attractiveness were much greater for a more photorealistic skin texture than for a less photorealistic skin texture.

6.1. Design principles for CG animators

The findings of this study suggest the following design principles. To design attractive, human-looking faces that are not eerie, use high polygon counts with smoothing and nearly ideal facial proportions. It may be safer to use a less photorealistic texture unless human photorealism is required. When using a human photorealistic texture, ensure the proportions of the CG face are within human norms. And finally, to prevent eeriness, avoid mismatches in the degree of human likeness of CG elements.

6.2. Limitations

A limitation of this study is the use of one-item measures for the reassuring–eerie and nonhuman–human scales in Study I and for the reassuring–eerie, artificial–natural, and ugly–attractive scales in Study IV. One-item measures preclude tests of reliability or construct validity. For that reason, we are currently developing an eeriness index, because none exists to date. Powers and Kiesler (2006) have developed a human likeness index whose items were adapted to a semantic differential scale by Bartneck (2008). However, in Studies I and IV the sheer quantity of stimuli that participants would be required to rate with multiple-item indices would raise concerns about fatigue effects and attrition. The number of stimuli in this study already necessitated the use of different participants in each experiment, which prevented the assessment of experimental crossover interaction.

Another limitation is that manipulations were only applied to one base model representing a 30-year-old male. Repeating Studies I through IV with several base models could rule out the possibility that effects were driven by the particular features of the model used. It would be worthwhile to use models of youths and older adults, and also female base models, because some preferences are specific to age and gender (Grammer & Thornhill, 1994; Henss, 1991). The generalization of the results to the broader US population or to other cultures is also constrained by the demographic homogeneity of the participants (i.e., undergraduates and recent graduates of a nine-campus Midwestern university). An additional concern is that Maya automatically determined the changes in the level of detail of the CG faces, including the placement of visual landmarks, such as lines and edges between polygons. This can produce unpleasant-looking results. When a CG animator increases or reduces a model’s level of detail, artistic judgment is involved. Some of the results of this study could be influenced by the automated design process.

Finally, the findings of this study only apply to still CG faces. They may not apply to animated faces or to still or animated bodies. If not well-animated, even nearly perfect human CG characters can look uncanny when moving (Chaminade et al., 2007).

NIH-PA Author Manuscript
NIH-PA Author Manuscript

6.3. Future work

Many CG animators are pursuing the holy grail of photorealistic human appearance–a kind of virtual 3D Turing test, *sans* the interactivity. Future studies on the relation between CG and the uncanny should also focus on temporal dynamics, such as motion quality, timing, and contingency during interaction.

7. Conclusion

Contrary to the predictions of Mori's graph of the uncanny valley, a CG face is not necessarily eeriest when it looks nearly human. Even abstract faces can look eerie, if they contain facets that seem unintended or arbitrary (Hanson, 2006; MacDorman & Ishiguro, 2006). Nevertheless, backing away from a photorealistic texture somewhat can decrease eeriness in CG faces with facial proportions that deviate from human norms. Adding polygons and smoothing to a face with a photorealistic or bronze texture increases perceived human likeness, but automatically adding lines to a line-textured face can increase eeriness after a point.

Distorting facial proportions causes more detailed faces to be rated eeriest. In particular, with a 50% enlargement of the eyes, the eeriness of a CG face with a photorealistic human texture will increase much more than the eeriness of a bronze-textured face. Also, as human likeness increases, the facial proportions perceived as best-looking approach the baseline proportions. These proportions may be closer to the human ideal. Thus, a photorealistic texture may be crucial in eliciting normative expectations concerning human proportions, and faces violating these expectations appear eerie. These results suggest potential neuroimaging studies that explore the interaction effect of facial texture and proportions on brain regions where heightened activation is associated with perceived eeriness.

Acknowledgments

We are grateful to Sandosh Vasudevan for developing and maintaining the website and tools used in the studies, Mike Bentley, Maria Britain, Charles R. Goodlett, Wade Mitchell, Himalaya Patel, Ryan Wing, and Thomas Wiggins for their insightful comments, and Thomas Busey and Alan C. Roberts for suggestions that inspired this work.

References

Andrews TJ, Schluppeck D, Homfray D, Matthews P, Blakemore C. Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. Neuroimage. 2002; 17:890–901. [PubMed: 12377163]

Atkinson AP, Heberlein AS, Adolphs R. Spared ability to recognise fear from static and moving whole-body cues following bilateral amygdala damage. Neuropsychologia. 2007; 45(12):2772–2782. [PubMed: 17561172]

Bailenson JN, Swinth KR, Hoyt CL, Persky S, Dimov A, Blascovich J. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. Presence: Teleoperators and Virtual Environments. 2005; 14(4):379–393.

Barbeau EJ, Taylor MJ, Regis J, Marquis P, Chauvel P, Liegeois-Chauvel C. Spatio temporal dynamics of face recognition. Cerebral Cortex. 2008; 18(5):997–1009. [PubMed: 17716990]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Bartneck, C. Who like androids more: Japanese or US Americans?. 17th IEEE international symposium on robot and human interactive communication (ROMAN); Munich: IEEE; 2008. p. 553-557.

Becker, E. The denial of death. New York: Free Press; 1973.

Burgoon JK, Bonito JA, Bengtsson B, Cederberg C, Lundeberg M, Allspach L. Interactivity in human–computer interaction: A study of credibility, understanding, and influence. Computers in Human Behavior. 2000; 16(6):553–574.

Canemaker, J. A part-human, part-cartoon species. The New York Times; 2004.

Carmel D, Bentin S. Domain specificity versus expertise: Factors influencing distinct processing of faces. Cognition. 2002; 83:1–29. [PubMed: 11814484]

Carr L, Iacoboni M, Dubeau M, Mazziotta J, Lenzi G. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. Proceedings of the National Academy of Sciences. 2003; 100(9):5497–5502.

Chaminade TD, Hodgin JK, Kawato M. Anthropomorphism influences perception of computer-animated characters' actions. Social Cognitive and Affective Neuroscience. 2007; 2(3):206–216. [PubMed: 18985142]

Cole J. Empathy needs a face. Journal of Consciousness Studies. 2001; 8:51–68.

Conway CA, Jones BC, DeBruine LM, Little A. Evidence for adaptive design in human gaze preference. Proceedings of the Royal Society: Biological Sciences. 2008; 275(1630):63–69. [PubMed: 17986435]

Costa M, Corassa L. Aesthetic phenomena as supernormal stimuli: The case of eye, lip and lower-face size and roundness in artistic portraits. Perception. 2006; 35(2):229–246. [PubMed: 16583768]

Cunningham MR. Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. Journal of Personality and Social Psychology. 1986; 50(5): 925–935.

Cunningham MR, Roberts AR, Barbee AP, Druen PB, Wu CH. "Their ideas of beauty are on the whole the same as ours?" Consistency and variability in cross-cultural perception of female physical attractiveness. Journal of Personality and Social Psychology. 1995; 68(2):261–279.

Davis M, Whalen PJ. The amygdala: Vigilance and emotion. Molecular Psychiatry. 2001; 6(1):13–34. [PubMed: 11244481]

Del Cul A, Baillet S, Dehaene S. Brain dynamics underlying the nonlinear threshold for access to consciousness. PLoS Biology. 2007; 5(10):2408–2423.

Diamond R, Carey S. Why faces are and are not special: An effect of expertise. Journal of Experimental Psychology: General. 1986; 115(2):107–117. [PubMed: 2940312]

Downing PE, Jiang Y, Shuman M, Kanwisher N. A cortical area selective for visual processing of the human body. Science. 2001; 293:2470–2473. [PubMed: 11577239]

Ekman P, Friesen WV. A new pan-cultural facial expression of emotion. Motivation and Emotion. 1986; 10(2):159–168.

Emery NJ. The eyes have it: The neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews. 2000; 24:581–604. [PubMed: 10940436]

Etcoff, N. Survival of the prettiest: The science of beauty. New York: Doubleday; 1999.

Farah MJ, Rabinowitz C, Quinn GE, Liu GT. Early commitment of neural substrates for face recognition. Cognitive Neuropsychology. 2000; 17:117–123. [PubMed: 20945175]

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain. 1996; 119:593–609. [PubMed: 8800951]

Gauthier I, Logothetis NK. Is face recognition not so unique after all? Cognitive Neuropsychology. 2000; 17:125–142. [PubMed: 20945176]

Gazzola V, Rizzolatti G, Wicker B, Keysers C. The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. NeuroImage. 2007; 35(4):1674–1684. [PubMed: 17395490]

Geller T. Overcoming the uncanny valley. IEEE Computer Graphics and Applications. 2008; 28(4): 11–17. [PubMed: 18663813]

Giddon DB, Sconzo R, Kinchen JA, Evans CA. Quantitative comparison of computerized discrete and animated profile preferences. The Angle Orthodontist. 1996; 66(6):441–448. [PubMed: 8974180]

Gong L. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. Computers in Human Behavior. 2008; 24(4): 1494–1509.

Gouskos, C. The depths of the uncanny valley. Game Spot. 2006. Available from http://www.gamespot.com/features/6153667/index.html

Grammer K, Thornhill R. Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. Journal of Comparative Psychology. 1994; 108(3):233–243. [PubMed: 7924253]

Green RD, MacDorman KF, Ho C-C, Vasudevan SK. Sensitivity to the proportions of faces that vary in human likeness. Computers in Human Behavior. 2008; 24(5):2456–2474.

Grezes J, Armony J, Rowe J, Passingham R. Activations related to "mirror" and "canonical" neurones in the human brain: An fmri study. Neuroimage. 2003; 18:928–937. [PubMed: 12725768]

Grill-Spector K, Knouf N, Kanwisher N. The fusiform face area subserves face perception, not generic within-category identification. Nature Neuroscience. 2004; 7(5):555–562.

Haidt J, McCauley C, Rozin P. Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. Personality and Individual Differences. 1993; 16:701–713.

Hanson, D. Exploring the aesthetic range of humanoid robots. Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science; 2006. p. 16-20.

Hasson U, Hendler T, Ben Bashat D, Malach R. Vase or face? A neural correlate of shape-selective grouping processes in the human brain. Journal of Cognitive Neuroscience. 2001; 13:744–753. [PubMed: 11564319]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001; 293:2425–2430. [PubMed: 11577229]

Henss R. Perceiving age and attractiveness in facial photographs. Journal of Applied Social Psychology. 1991; 21:933–946.

Ho, C-C.; MacDorman, KF.; Pramono, ZAD. Human emotion and the uncanny valley: A GLM, MDS, and isomap analysis of robot video ratings. ACM/IEEE international conference on human–robot interaction; 2008.

Jabbi M, Bastiaansen J, Keysers C. A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. PLoS ONE. 2008; 3(8):e2939. [PubMed: 18698355]

James LR, Demaree RG, Wolf G. r_{wg}: An assessment of within-group interrater agreement. Journal of Applied Psychology. 1993; 78(2):306–309.

Jones BC, Little AC, Perrett DI. When facial attractiveness is only skin deep. Perception. 2004; 33(5): 569–576. [PubMed: 15250662]

Jones D. Sexual selection, physical attractiveness, and facial neoteny: Crosscultural evidence and implications. Current Anthropology. 1995; 36(5):723–748.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience. 1997; 17:4302–4311. [PubMed: 9151747]

Keysers C, Gazzola V. Integrating simulation and theory of mind: From self to social cognition. Trends in Cognitive Sciences. 2007; 11(5):194–196. [PubMed: 17344090]

Krach S, Hegel F, Wrede B, Sagerer G, Binkofski F, Kircher T. Can machines think? Interaction and perspective taking with robots investigated via fmri. PLoS ONE. 2008; 3(7):e2597. [PubMed: 18612463]

Langlois JH, Kalakanis L, Rubenstein AJ, Larson A, Hallam M, Smoot M. Maxims or myths of beauty? A meta-analytic and theoretical review. Psychological Bulletin. 2000; 126(3):390–423. [PubMed: 10825783]

Langlois JH, Roggman LA. Attractive faces are only average. Psychological Science. 1990; 1(2):115–121.

Langlois JH, Roggman LA, Casey RJ, Ritter JM, Rieser-Danner LA, Jenkins VY. Infant preferences for attractive faces: Rudiments of a stereotype. Developmental Psychology. 1987; 23(3):363–369.

Law Smith MJ, Perrett DI, Jones BC, Cornwell RE, Moore FR, Feinberg DR, et al. Facial appearance is a cue to oestrogen levels in women. Proceedings of the Royal Society: Biological Sciences. 2006; 273(1583):135–140. [PubMed: 16555779]

Little AC, Jones BC. Evidence against perceptual bias views for symmetry preferences in human faces. Proceedings of the Royal Society of London: Biological Sciences. 2003; 270:1759–1763.

MacDorman, KF. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. ICCS/CogSci- 2006 long symposium: Toward social mechanisms of android science; Vancouver, Canada. 2006.

MacDorman KF, Ishiguro H. The uncanny advantage of using androids in cognitive and social science research. Interaction Studies. 2006; 7(3):297–337.

MacDorman, KF.; Minato, T.; Shimada, M.; Itakura, S.; Cowley, SJ.; Ishiguro, H. Assessing human likeness by eye contact in an android testbed. Proceedings of the XXVII annual meeting of the cognitive science society; Stresa, Italy. 2005.

MacDorman KF, Vasudevan SK, Ho CC. Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. AI and Society. 2009; 23(4):485–510.

MacGillivray, C. How psychophysical perception of motion and image relates to animation practice. Imaging and Visualization, International Conference on Computer Graphics; 2007. p. 81-88.

MacMillan, D. Navigating the uncanny valley. Business Week. 2007. Available from http://businessweek.com/innovate/content/aug2007/id20070817_955317.htm

Manning JT, Gage AR, Diver MJ, Scutt D, Fraser WD. Short-term changes in asymmetry and hormones in men. Evolution and Human Behavior. 2002; 23:95–102.

Manning JT, Pickup LJ. Symmetry and performance in middle distance runners. International Journal of Sports Medicine. 1998:205–209. [PubMed: 9630027]

Manning JT, Scutt D, Lewis-Jones DI. Developmental stability, ejaculate size and sperm quality in men. Evolution and Human Behavior. 1998:273–282.

Mori M. Bukimi no tani [The uncanny valley]. Energy. 1970; 7(4):33–35.

Norton KI, Olds TS, Olive S, Dank S. Ken and Barbie at life size. Sex Roles. 1996; 43(3–4):287–294.

Nowak KL, Biocca F. The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments. Presence: Teleoperators and Virtual Environments. 2003; 12(5):481–494.

Ohman, A. Fear and anxiety: Evolutionary, cognitive, and clinical perspectives. In: Lewis, M.; Haviland-Jones, JM., editors. Handbook of emotions. New York: The Guliford Press; 2000. p. 573-593.

Olson IR, Marshuetz C. Facial attractiveness is appraised in a glance. Emotion. 2005; 5(4):498–502. [PubMed: 16366753]

Pollick, FE. In search of the uncanny valley. In: Grammer, K.; Juett, A., editors. Analog communication: Evolution, brain mechanisms, dynamics, simulation. Cambridge, MA: MIT Press, The Vienna Series in Theoretical Biology; in press

Powers, A.; Kiesler, S. The advisor robot: Tracing people's mental model from a robot's physical attributes. Proceedings of the First ACM SIGCHI/SIGART Conference on Human-robot interaction; 2006.

Preston S, de Waal F. Empathy: Its ultimate and proximate bases. Behavioral and Brain Sciences. 2002; 25:1–72. [PubMed: 12625087]

Pyszczynski T, Greenberg J, Solomon S. A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory. Psychological Review. 1999; 106(4):835–845. [PubMed: 10560330]

Ramey, CH. The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. Proceedings of views of the uncanny valley workshop: IEEE-RAS international conference on humanoid robots; Tsukuba, Japan. 2005. p. 8-13.

Rhodes G, Proffitt F, Grady JM, Sumich A. Facial symmetry and the perception of beauty. Psychonomic Bulletin & Review. 1998; 5(4):659–669.

Rhodes G, Tremewan T. Averageness, exaggeration, and facial attractiveness. Psychological Science. 1996; 7(2):105–110.

Rhodes G, Yoshikawa S, Clark A, Lee K, McKay R, Akamatsu S. Attractiveness of facial averageness and symmetry in non-Western cultures: In search of biologically based standards of beauty. Perception. 2001; 30(5):611–625. [PubMed: 11430245]

Rhodes, G.; Zebrowitz, LA., editors. Facial attractiveness: Evolutionary cognitive and social perspectives. Westport, CT: Ablex; 2002.

Rozin P, Fallon AE. A perspective on disgust. Psychological Review. 1987; 94:23–41. [PubMed: 3823304]

Salvia J, Sheare JB, Algozzine B. Facial attractiveness and personalsocial development. Journal of Abnormal Child Psychology. 1975; 3(3):171–178. [PubMed: 1214029]

Seyama J, Nagayama RS. The uncanny valley: The effect of realism on the impression of artificial human faces. Presence: Teleoperators and Virtual Environments. 2007; 16(4):337–351.

Soler C, Nunez M, Gutierrez R, Nunez J, Medina P, Sancho M, et al. Facial attractiveness in men provides clues to semen quality. Evolution and Human Behavior. 2003; 24:199–207.

Solomon S, Greenberg J, Pyszczynski T. Tales from the crypt: The role of death in life. Zygon: Journal of Religion and Science. 1998; 33:9–43.

Spiridon M, Kanwisher N. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron. 2002; 35:1157–1165. [PubMed: 12354404]

Tai YF, Scherfler C, Brooks DJ, Sawamoto N, Castiello U. The human premotor cortex is mirror only for biological actions. Current Biology. 2004; 14:117–120. [PubMed: 14738732]

Thornhill R, Gangestad SW. Facial attractiveness. Trends in Cognitive Sciences. 1999; 3:452–460. [PubMed: 10562724]

Thornhill R, Gangstead S. Human facial beauty: Averageness, symmetry, and parasite resistance. Human Nature. 1993; 4:237–269. [PubMed: 24214366]

Tovée MJ, Hancock PJB, Mahmoodi S, Singleton BRR, Cornelissen PL. Human female attractiveness: Waveform analysis of body shape. Proceedings of the Royal Society of London Series B, Biological Sciences. 2002; 269:2205–2213.

Turkle S. Authenticity in the age of digital companions. Interaction Studies. 2007; 8(3):501–517.

Turkle S, Taggart W, Kidd CD, Daste O. Relational artifacts with children and elders: The complexities of cybercompanionship. Connection Science. 2006; 18(4):347–361.

Vinayagamoorthy, V.; Steed, A.; Slater, M. Toward social mechanisms of android science: A COGSCI 2005 workshop. Stresa, Italy: 2005. Building characters: Lessons drawn from virtual environments; p. 119-126.

Westen D, Blagov P, Harenski K, Kilts C, Hamann S. Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 US Presidential Election. Journal of Cognitive Neuroscience. 2006; 18(11):1947–1958. [PubMed: 17069484]

Willis J, Todorov A. First impressions: Making up your mind after a 100- ms exposure to a face. Psychological Science. 2006; 17(7):592–598. [PubMed: 16866745]

Zacharek S. Pirates of the Caribbean: At world's end. Salon. 2007

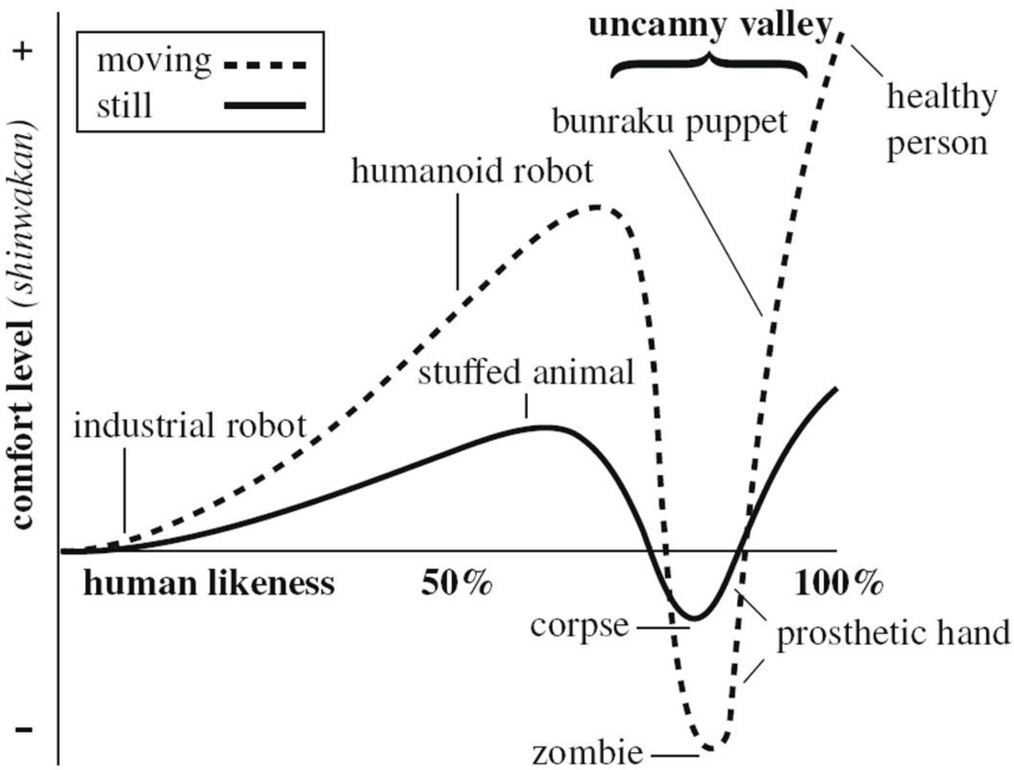

Fig. 1.
Masahiro Mori proposed a relation between human likeness and *shinwakan*, which may be roughly translated as rapport or comfort level: more human-looking robots are perceived as more agreeable until we get to robots that look so nearly human that subtle flaws make them look creepy. This dip in their evaluation is the uncanny valley. The valley, Mori argued, would be deepened by movement. The term *uncanny valley* is now commonly applied to animated characters in films and video games.

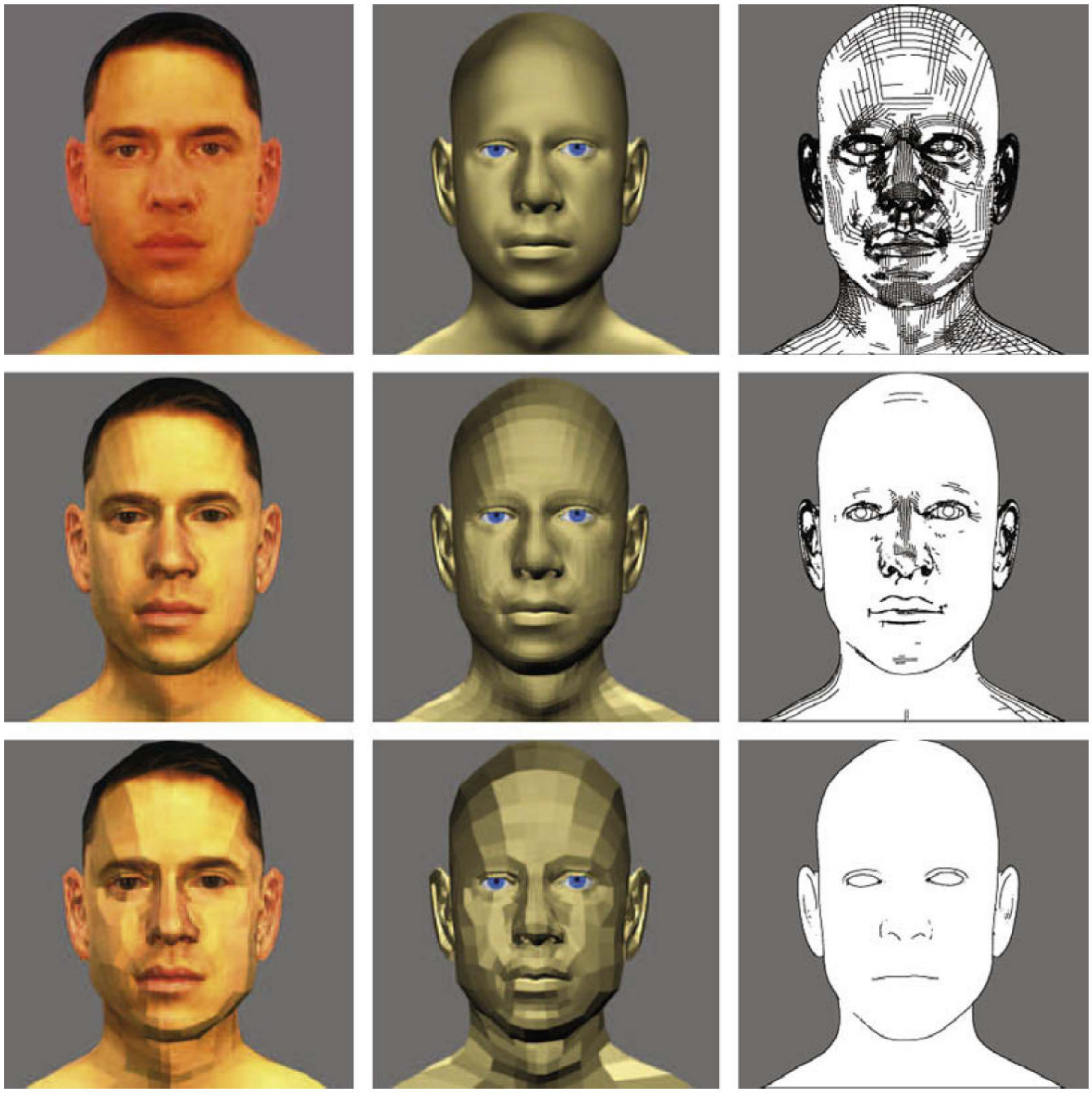

Fig. 2.
Photorealistic, bronze, and line texture models at varying levels of detail.

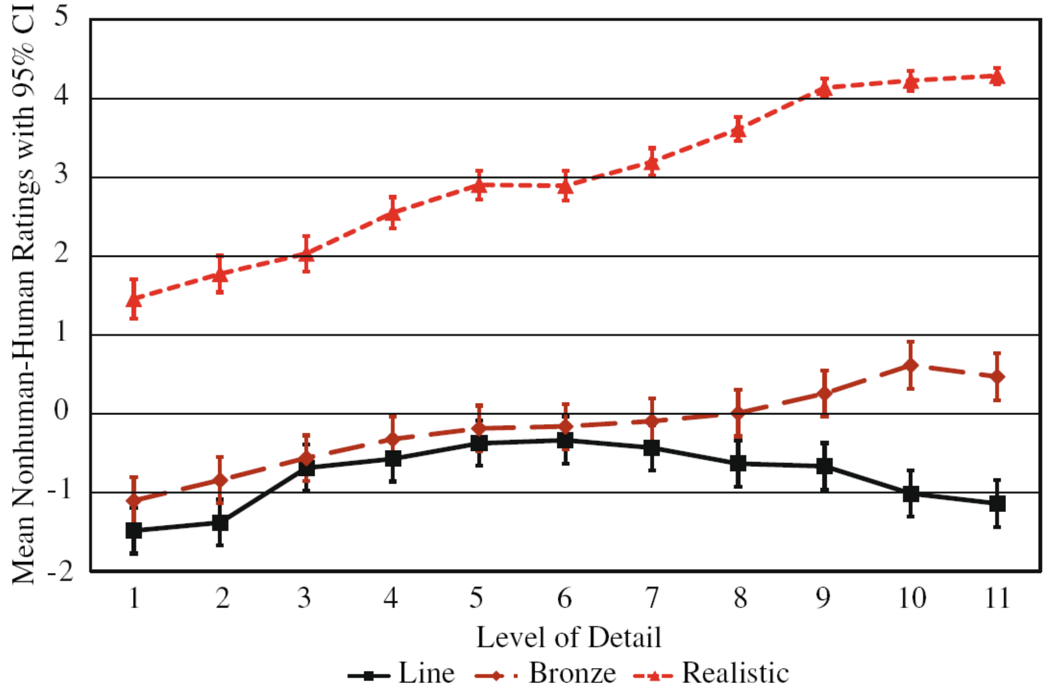

Fig. 3.
Mean human likeness by texture and detail.

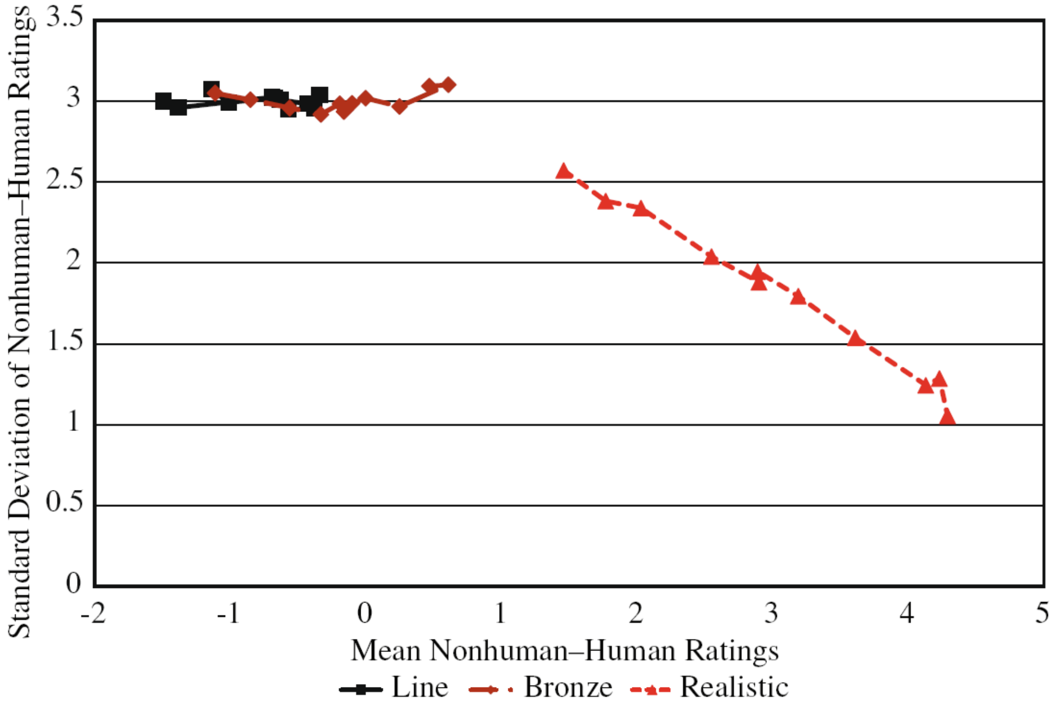

Fig. 4.
Standard deviation in human likeness by mean.

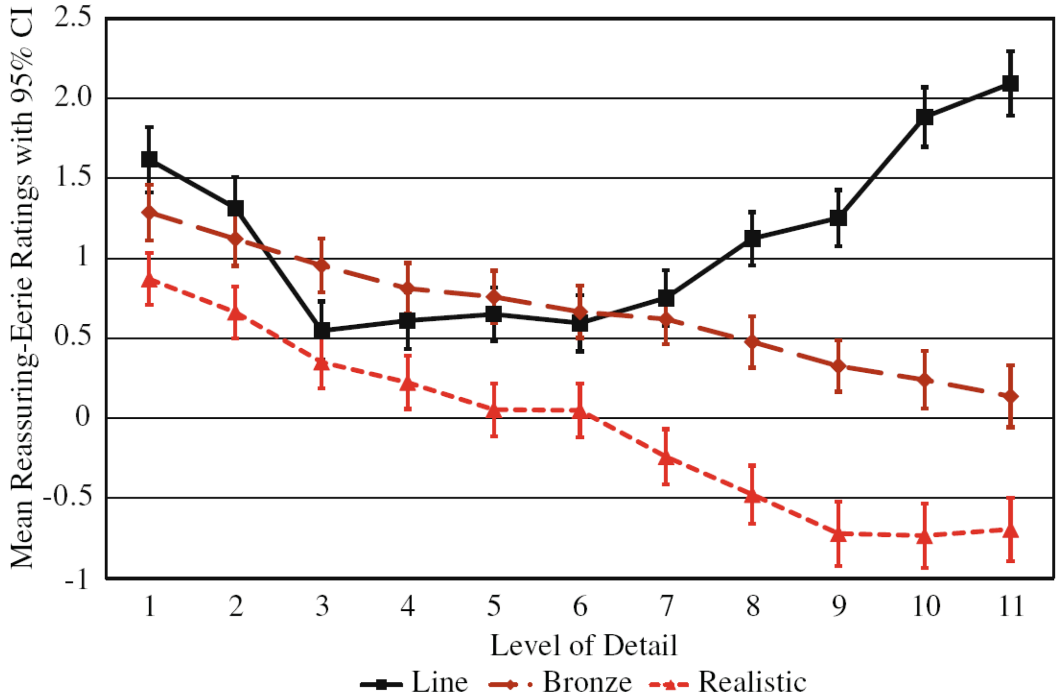

Fig. 5.

Mean eeriness by texture and detail.

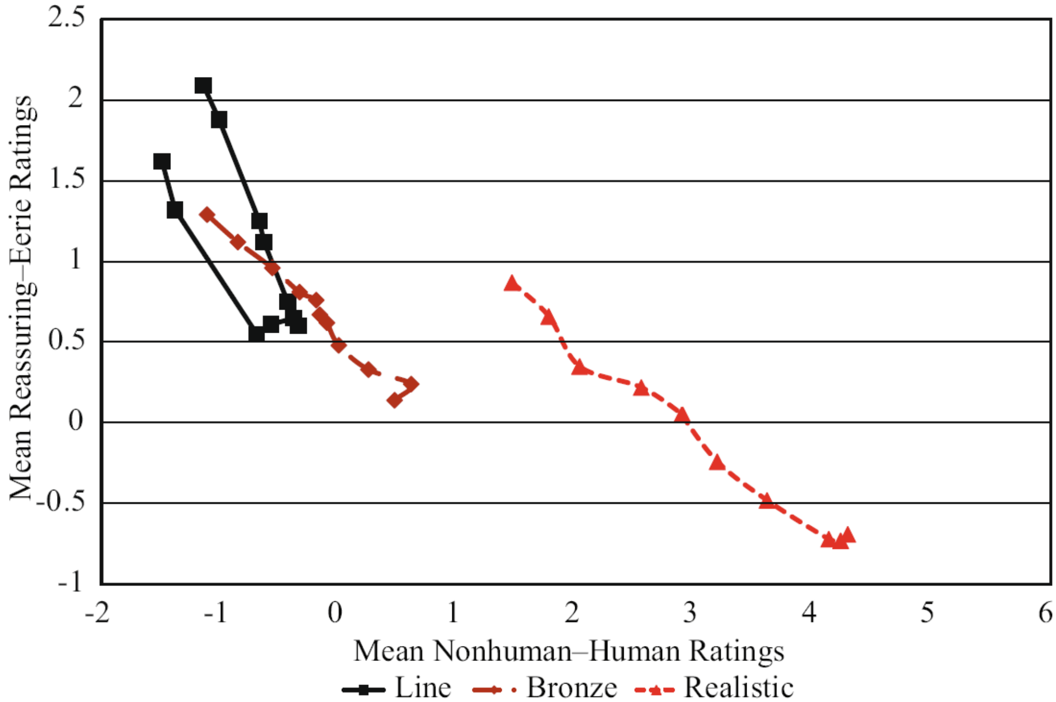

Fig. 6.
Mean eeriness by mean human likeness.

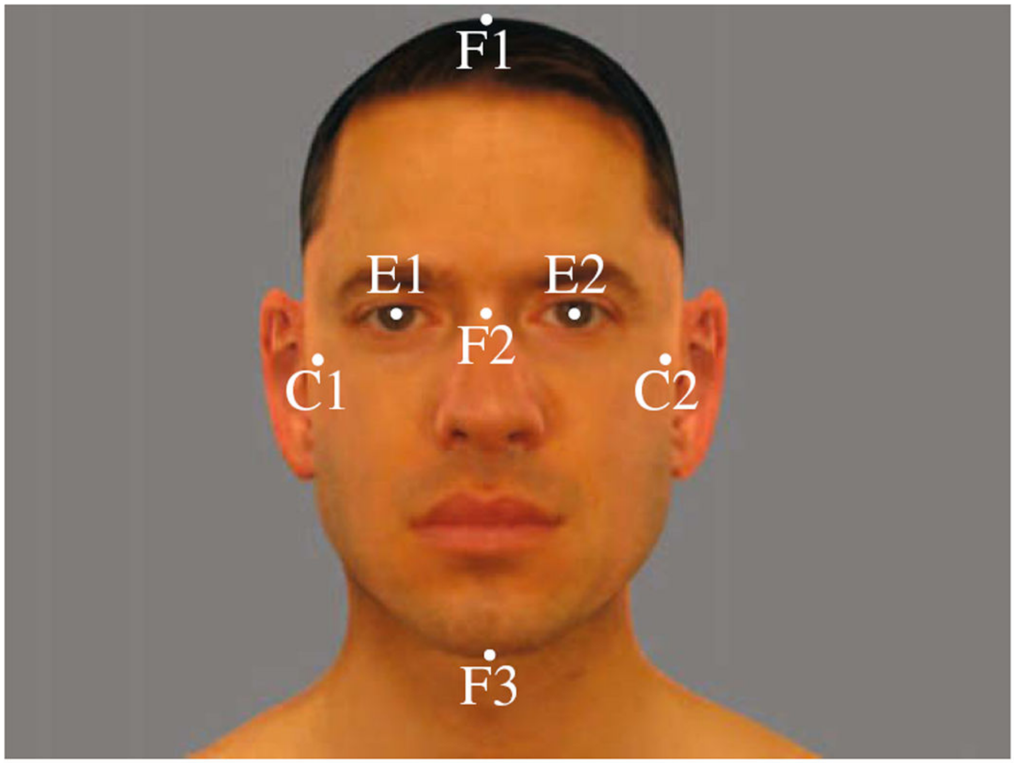

Fig. 7.
Points used in measuring facial dimensions.

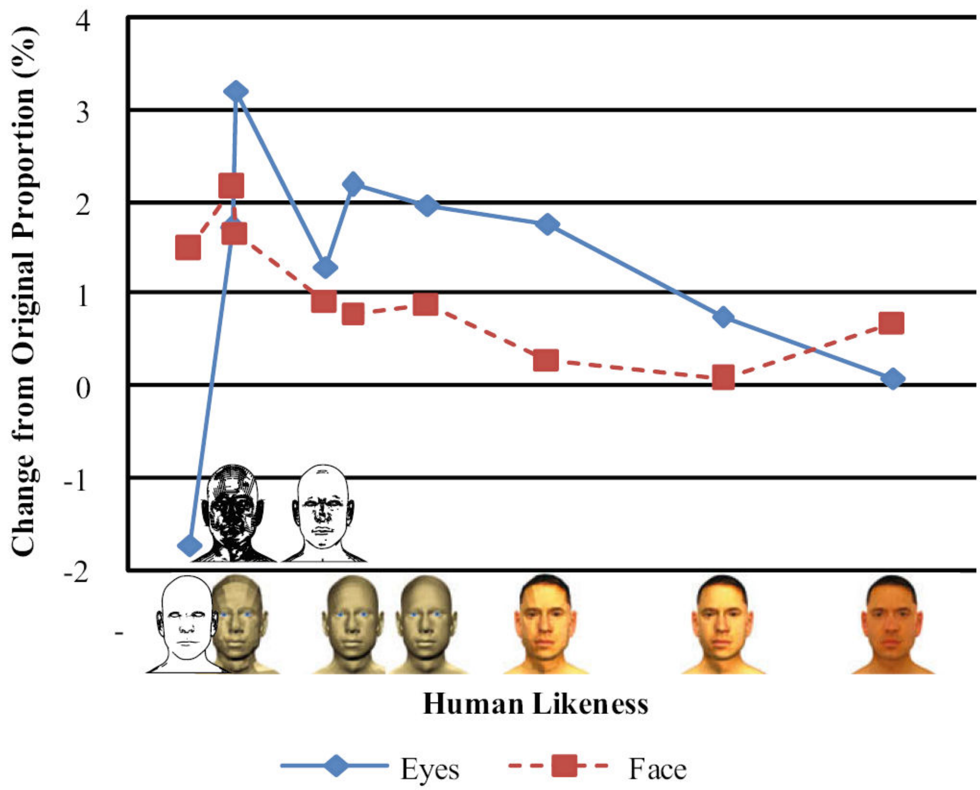

Fig. 8.
Mean best point by mean human likeness.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Fig. 9.
Facial extremes: ±10% eye separation and face height.

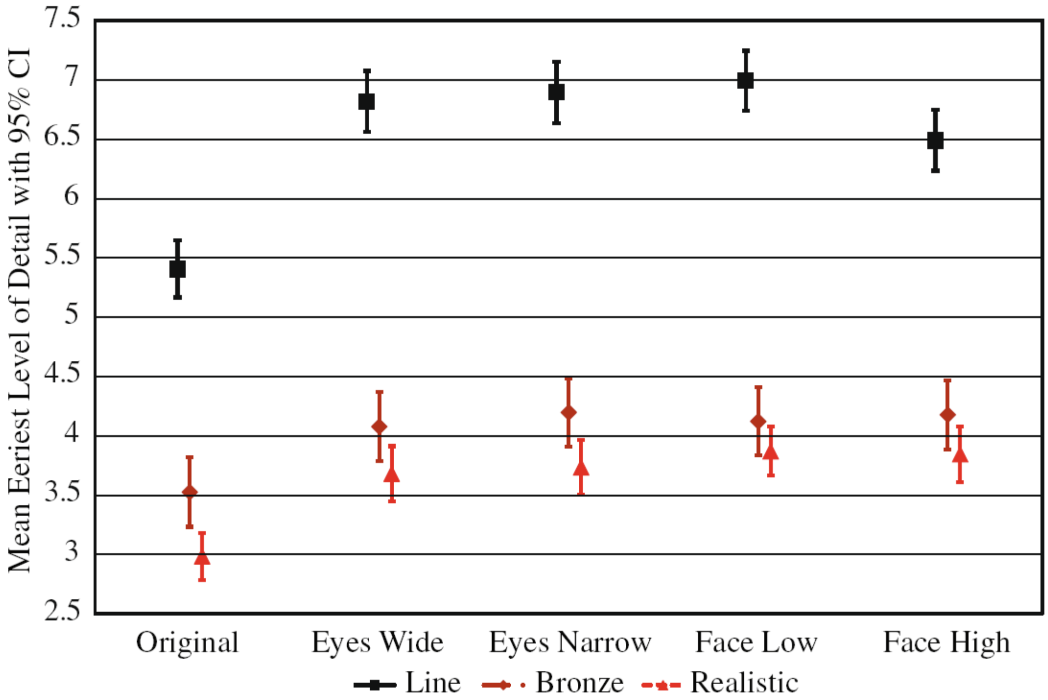

Fig. 10.
Mean eeriest point by texture and facial position.

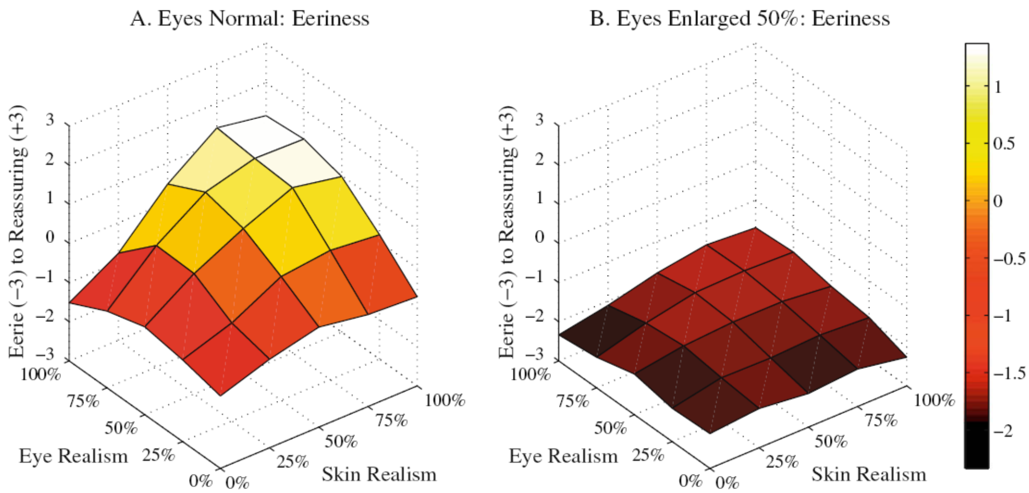

Fig. 11.
Mean eeriness by degree of eye and skin photorealism.

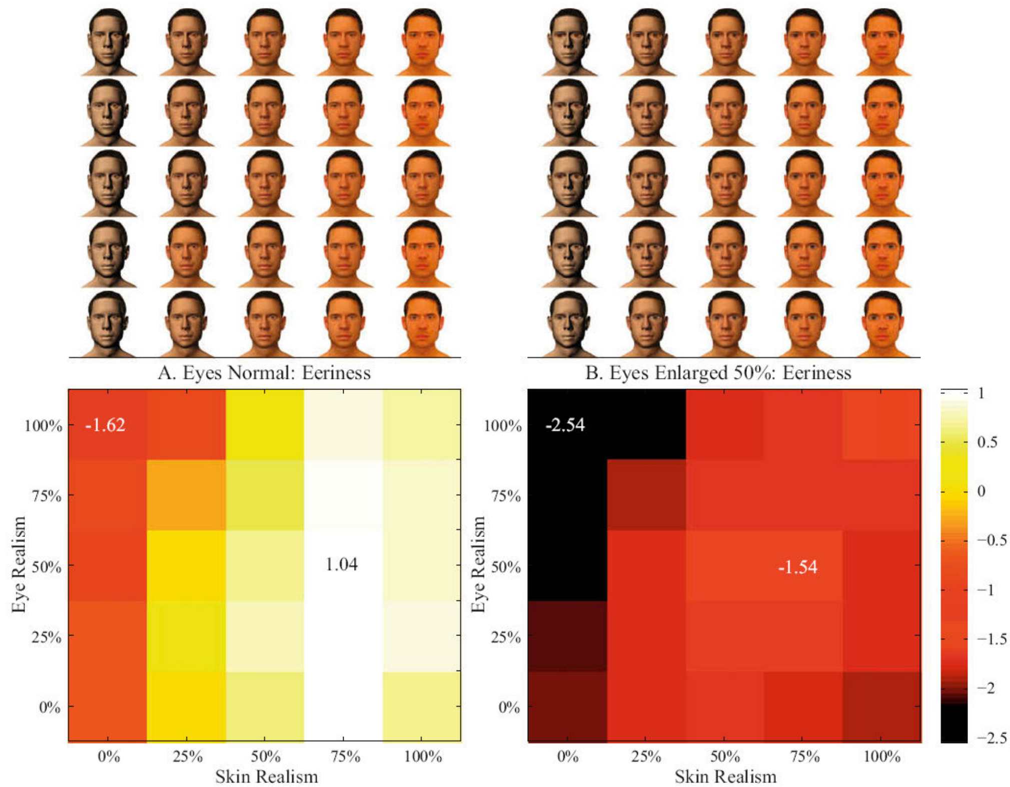

Fig. 12.
Mean eeriness by degree of eye and skin photorealism.

Table 1

One-way ANOVA.

Texture	**Eeriest point**		**Eeriness sensitivity**	
	$F(4,4254)$	r	$F(4,4254)$	r
Line	15.29***	.11	0.43	.01
Bronze	5.82***	.07	2.44*	.04
Photorealistic	10.76***	.10	4.83**	.06

* $p < .05$.

** $p < .01$.

*** $p < .001$.